

DOCUMENT RESUME

ED 321 514

FL 017 667

AUTHOR Derbyshire, Desmond C., Ed.
 TITLE 1987 Work Papers of the Summer Institute of Linguistics, University of North Dakota Session. Volume 31.
 INSTITUTION Summer Inst. of Linguistics, Grand Forks, N. Dak.
 PUB DATE 87
 NOTE 130p.
 AVAILABLE FROM SIL/UND Work Papers, c/o International Linguistic Center-Bookroom, 7500 West Camp Wisdom Road, Dallas, TX 75236.
 PUB TYPE Collected Works - General (020) -- Reports - Research/Technical (143)
 EDRS PRICE MF01/PC06 Plus Postage.
 DESCRIPTORS Dictionaries; Foreign Countries; Grammar; *Greek; Guarani; *Language Patterns; Language Research; *Linguistic Theory; *Pragmatics; *Semantics; Structural Analysis (Linguistics); *Syllables; Uncommonly Taught Languages
 IDENTIFIERS Brazil; Clauses; Hixkaryana; *Seri; Xavante

ABSTRACT

Four working papers in linguistics are presented. "Basic Configurations of Pragmatic Structuring" (Robert A. Dooley) presents evidence for basic or unmarked pragmatic configurations in three Brazilian languages: Topic-Comment in Mbya Guarani, Comment-Topic in Hixkaryana, and Focus-Presupposition in Xavante. "The Syllable Structure of Seri" (Stephen A. Marlett) contains a detailed discussion of issues relevant to syllable structure in the Seri language, including the sonority hierarchy, the internal structure of syllables, representation of long vowels and round consonants, the relationship of stress to syllabification, and the distinction between the syllable structure of inflected and uninflected words. "Extraction from Complement Clauses in Koine Greek" (Stephen A. Marlett) examines one constraint on word order, and shows how it affects the understanding of two verses (from the New Testament/1 Corinthians 14:12 and 1 John 5:16) that seem to be counterexamples to it. "From Meaning to Text: Semantic Representation in the Meaning-Text Linguistic Theory and a New Type of Monolingual Dictionary" (Igor A. Melcuk) expounds a different approach to natural language, the Meaning-Text Theory, focusing on the theory's general structure and a new type of lexicon. (MSE)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *



WORK PAPERS

VOLUME XXXI

1987

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

R. Dooley

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.
☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.



SUMMER INSTITUTE OF LINGUISTICS
UNIVERSITY OF NORTH DAKOTA SESSION

FL 017667

1987
WORK PAPERS
OF THE
SUMMER INSTITUTE OF LINGUISTICS
University of North Dakota
Session

Volume 31

Editor: Desmond C. Derbyshire



These are working papers and should not be cited without referring to their preliminary nature.

Orders and correspondence should be sent to:

SIL/UND Work Papers
c/o International Linguistic Center-Bookroom
7500 West Camp Wisdom Road, Dallas, TX 75236

PREFACE

Three of the papers in this volume present provisional results of research by two members of the faculty of the Summer Institute of Linguistics at the University of North Dakota, Steve Marlett and Bob Dooley. The other is a paper presented in the 1986 SIL-UND colloquium series by Igor Mel'čuk, who was a visiting lecturer that summer.

I would like to thank Kathie Dooley for handling the composition and production, ably assisted by Nathan Dooley. My thanks also to Betty Brown, Hu Matthews and Diana Weber for their editing help.

D.C.D.

TABLE OF CONTENTS

Preface	111
Basic Configurations of Pragmatic Structuring Robert A. Dooley	1
The Syllable Structure of Seri Stephen A. Marlett	29
Extraction from Complement Clauses in Koine Greek Stephen A. Marlett	65
From Meaning to Text: Semantic Representation in the Meaning-Text Linguistic Theory and a New Type of Monolingual Dictionary Igor A. Mel'čuk	73

BASIC CONFIGURATIONS OF PRAGMATIC STRUCTURING

Robert A. Dooley

- 1 Introduction
- 2 Pragmatic functions and configurations
- 3 Topic-Comment in Mbyá Guaraní
- 4 Comment-Topic in Hixkaryana
- 5 Focus-Presupposition in Xavante
- 6 "V-only" languages
- 7 Toward universals of pragmatic structuring
- 8 Conclusion

1 Introduction

In this paper, I present evidence for basic or unmarked pragmatic configurations in three languages of Brazil: Topic-Comment in Mbyá Guaraní, Comment-Topic in Hixkaryana, and Focus-Presupposition in Xavante. In doing this, I present a framework for the description of pragmatic functions and configurations, and discuss as well certain possible universals which can be stated in this framework.

In language, as in many other human activities, when a set of choices is presented often enough, there is often one option which comes to function as the unmarked choice; that is, it is chosen in the absence of factors that would strongly indicate another. Thus, many languages have a "basic word order" stated in terms of grammatical relations such as S(ubject), O(bject), and V(erb). In English, for example, the statement that the basic order of these elements is SVO is fundamental, at some level, for the description of sentence syntax.

Alongside the kind of structure that is stated in terms of grammatical relations, there is also a kind that is stated in terms of pragmatic functions. Hockett, following Sapir (1921), was the first to use the terms "Topic" and "Comment" in pointing out that English has a basic configuration of Topic-Comment. He described it as a "favorite sentence-type" for the language, which usually, but not always, corresponds to the grammatical constituent structure of subject and predicate (Hockett 1958:201ff; cf. also Hockett 1966:23).¹

The Topic-Comment configuration is apparently common in languages, so much so that basic characteristics of human information processing have been adduced to account for its preponderance. Keenan, in attempting to explain the scarcity of languages with subject-final word order, cites prototypical topic-like properties of subjects, and then comments that "topics in general come first because they determine the relevance of what is said for the addressee" (1978:305). Similarly, Mallinson and Blake (1981), in stating as a general cross-linguistic principle that "more topical material tends to come nearer to the

beginning of the clause (to the left) than non-topical material" (151), claim that this ordering is found "in mediums other than language. In mime and dance this is true, and it is true of at least some types of visual display ... Normally comments do not make sense without topics" (156).

If human cognition were no more complex than that, a language universal could probably be stated in more or less the following terms:

- (1) "The ordering Topic-Comment is the only basic pragmatic configuration in human languages; other configurations only occur as marked choices."

But statement 1 is not the case. Hockett, in positing a weaker universal ("Every human language has a common clause type with a bipartite structure in which the constituents can reasonably be termed 'topic' and 'comment'"), goes on to add that "the order of the constituents varies. Typically in Chinese, Japanese, Korean, English, and many other languages, one first mentions something that one is going to talk about, and then says something about it. In other languages, the most typical arrangement is for the comment, or part of it, to precede the topic" (1966:23). Similarly, Mallinson and Blake state that "it is not impossible for a topic to appear to the right of the comment. This happens some of the time in any language and it is not altogether surprising that a few languages regularly put the subject to the right, e.g. Malagasy, Tzotzil and Houailou" (1981:156). But neither Hockett nor Mallinson and Blake actually claim Comment-Topic as the basic ordering of these functions in any particular language; the basic OS order in the languages cited by Mallinson and Blake, for example, could possibly be explained in terms of a disassociation of Subject and Topic.

In this paper, I will be making a small exploration into the question: What variety is present in human languages in regard to a basic, unmarked pragmatic configuration? After presenting some preliminary notions (Sect. 2) and the evidence for Topic-Comment in Mbyá Guaraní, Comment-Topic in Hixkaryana, and Focus-Presupposition in Xavante (Sects. 3-5), I will discuss the problem of V-only languages (Sect. 6), and then some considerations regarding a universal framework for the study of pragmatic configurations (Sect. 7).

It should be admitted at the outset that the identification of a language's basic pragmatic configuration is fraught with many of the difficulties that plague investigations of basic orderings of grammatical relations: conflicting results between such criteria as frequency counts and simplicity of "movement" rules; the scarcity of sentences with a full complement of the elements under investigation; and the interaction of grammatical and pragmatic phenomena in the structuring of sentences (Prody 1982). Certain difficulties are perhaps more acute in establishing a basic pragmatic configuration: pragmatic functions such as Topic and Focus, in general, have even fewer generally agreed-upon distinguishing features in surface structure than S and O. On the other hand, the task at hand can be considered simpler in the

sense that the distinction between surface and underlying orders is less relevant to pragmatic functions than to grammatical relations, in the sense that the identification of pragmatic functions closely follows surface signals.

2 Pragmatic functions and configurations

In this paper I will be using a framework for pragmatic functions and configurations which was developed first for the description of Mbyá Guaraní (Dooley 1982). In conceptualizing the general notion of pragmatic function, I begin with a definition from Dik's **Functional grammar**: "By pragmatic functions we understand functions which specify the informational status of the constituents involved within the wider communicative setting in which they occur" (Dik 1978:128). I go beyond Dik, however, in the following: When an overall pragmatic configuration for a sentence is involved, especially one which is prominent in the language, I view the pragmatic functions as syntagmatic relations, parallel to subject and object in grammar, not merely as categories parallel to noun phrase. For example, I view Topic as a syntagmatic relation that only has meaning in a configuration such as Topic-Comment, and in syntagmatic relation with the Comment. In this paper, I refer to five pragmatic functions: Core, Topic, Setting, Presupposition, and Tail.

I use **Core** as a cover term to include the two common notions **Focus** (in a Focus-Presupposition configuration) and **Comment** (in a Topic-Comment configuration). As such, Core can be characterized as a pragmatic function comprising the part of a sentence which, in the context, has the highest informational value (cf. "Focus" in Dik 1978:130, 149ff and Comrie 1981:57). In many languages, the Core component of a sentence regularly receives the intonation center (Daneš 1967(1972):225f; Mallinson & Blake 1981:152), but this is not the case in some tonal languages (Watters 1979:138), nor even in non-tonal Hixkaryana, where "there is no special stress or emphasis on constituents" that are fronted for focus (Derbyshire 1985:146). However, I make the assumption that all pragmatic configurations will have a Core in some form as one of their pragmatic functions.

Topic is usually described as that entity which the sentence (or Comment/Core) is "about" (Dik 130, 141ff; Andrews 1985:77); it can perhaps better be described as that entity (when any such is indicated) whose "address" in the hearer's memory is the principal place at which the information in the sentence is to be attached or "filed" (cf. Reinhart 1982:24). This amounts to a cognitive explication of the notion of "aboutness". When such an address is explicitly named as Topic, the hearer must be able to find the address in his own cognitive inventory. For this reason, two properties of Topics follow from the above characterization: (i) Topics, as names of cognitive addresses, are prototypically nominals; and (ii) Topics, as hearer-recognizable addresses, are definite (Li & Thompson 1976:461). (It is understood that generic classes are definite in this sense.) In the sentence **Cats, I can't stand**, **cats** is the Topic and **I can't stand** is the Comment/Core.

I distinguish between "inner Topics" and "outer Topics": an **inner Topic** fills one of the nuclear grammatical relations in a clause, whereas an **outer Topic** is grammatically an adjunct or satellite to the clause. What I call outer Topic, Dik refers to as Theme (1978:130). But his characterization of Theme as presenting "a domain or universe of discourse with respect to which it is relevant to pronounce the following Predication" is, to me, indistinguishable from certain characterizations of Topic, such as Chafe's comment on Topics in Mandarin: "What the topics appear to do is to limit the applicability of the main predication to a certain restricted domain ... the topic sets a spatial, temporal, or individual framework within which the main predication holds" (1976:50). The following sentence, similar to one in Dik (1978:141), illustrates the difference between inner and outer Topics:

- (2) ---OTopic--- (-----ITopic----- ----Comment----)
 As for Paris, the Eiffel Tower is breathtaking.

The bracketing in 2, indicated by parentheses around the inner Topic and its Comment, indicates that this comprises the semantic scope of the outer Topic **as for Paris**. By means of the term Core, we can say that the Core of the outer Topic is itself structured into a Topic-Core configuration. Thus, different layers of pragmatic structuring are possible in a single sentence (Dooley 1982; Andrews 1985:80).

The pragmatic function **Setting** is like Topics in providing "a spatial or temporal framework within which the main predication holds" (see reference to Chafe above), but differs in that it is prototypically an adverbial rather than a nominal. That is, Settings and Topics have much in common functionally, but there are certain formal differences. In comparing English and Mandarin in this regard, Chafe mentions the sentence

- (3) ----Setting----- -----Grammatical nucleus-----
 In Dwinelle Hall people are always getting lost.

(my parsing, RAD), and then says that "Chinese would not require the **in**" (1976:51). In the terms of this paper, we could say that Chinese prefers outer Topics where English prefers Settings. Like outer Topics, Settings tend to be adjuncts or satellites to the grammatical nucleus of the clause. Typically, this is manifested in a separate intonation contour, or at least an intonation break, from the grammatical nucleus.

The pragmatic function **Presupposition** is describable as a propositional framework within which the filler of a certain "slot" is missing, that is, representable only as a variable. This pragmatic function occurs only in the Focus-Presupposition configuration, and in that configuration the missing content of the slot in question is furnished by the Focus component (Chomsky 1971:199ff). The characterization of Topic as a nominal and Presupposition as a propositional framework generally serves to distinguish these two pragmatic functions which occur with the Core. The content of a

Presupposition component "represents a situation with which the hearer is presumed to be familiar" (Andrews 1985:79). Typically, a Presupposition component consists of given information in the sense of information "which the speaker assumes to be in the consciousness of the addressee at the time of the utterance". Although it might include certain types of downgraded new information, that information is not being differentiated from given information by linguistic means (Chafe 1976:30,34). In line with this, I adopt a pragmatic rather than formal interpretation of "presupposition", as "assumptions the speaker makes about what the hearer is likely to accept without challenge" (Givón 1979:50). There may be many such presuppositions associated with a given sentence, but the pragmatic function Presupposition is a linguistically encoded unit, occurring syntagmatically with Focus. In the present paper, Focus-Presupposition as a pragmatic configuration refers to a binary division of a sentence in which these two components are continuous and occur in the order given. The cleft sentence 4 illustrates this configuration:

- (4) --Focus- Presupposition
 It's you that I love.

A fifth pragmatic function I will be referring to is what Dik calls **Tail**: "A constituent with Tail function presents, as an 'afterthought' to the Predication, information meant to clarify or modify (some constituent contained in) the Predication" (1978:130); "the Tail will characteristically be set off from the predication by means of a break in the intonation pattern" (153). The following example is also from Dik:

- (5) Grammatical nucleus ----Tail----
 He's a nice chap, your brother.

(I am using the term "grammatical nucleus" in place of Dik's "predication".)

In summary, then, of the five pragmatic functions that I have mentioned, three occur within the grammatical nucleus of a clause, and can accordingly be referred to as **nuclear pragmatic functions**: Core (including Comment and Focus), inner Topic, and Presupposition. Pragmatic configurations made up of nuclear pragmatic functions can be referred to as **nuclear pragmatic configurations**. It would be premature to claim that Core, inner Topic, and Presupposition constitute an exhaustive inventory of nuclear functions that occur in language. But it does appear to be the case that, crosslinguistically, there is a very small number of nuclear pragmatic functions and configurations. Andrews mentions three such "articulations": Topic-Comment, Focus-Presupposition, and Presentational (77). I would add two further ones: Comment-Topic and Focus-only. This last configuration would cover such sentence types as interjections (**Wow!**) and ideophones, but probably not types involving ellipsis (Q: **Who's speaking?** A: **John.**), since ellipsis can alternatively be analyzed as having other pragmatic functions (Topic or Presupposition) implicitly present. The Presentational or Existential

configuration is a special-purpose one to "introduce a new entity into the discourse...: **Once there was a king with three children**" (Andrews 1985:80). Perhaps it would be more accurate to say that the Presentational configuration introduces a new major entity. It may be possible to parse the Presentational configuration into inner Setting (a nuclear function corresponding to inner Topic) plus Core, often in that order (Givón 1983:34f).

- (6) Inner setting ----Core---
In the brush was a snake.

In a Presentational configuration, the Core contains a reference to the entity being introduced or brought into the scene.

With nuclear configurations, a certain amount of layered bracketing is possible. Thus, we might find Topic-(Focus-Presupposition) (Andrews 1985:80), Focus-(Topic-Comment), or even Focus-(Topic-(Topic-Comment)); all of these possibilities are illustrated for Mbyá Guaraní in Dooley (1982:317). It appears that only Comment and Presupposition can be restructured as embedded pragmatic configurations in this way. Accordingly, we might call Topic and Focus **simple** functions, and Comment and Presupposition **potentially complex** ones. Note that one form of Core is simple (i.e. Focus), and another is potentially complex (Comment). The difference between the simple and the potentially complex functions is probably a consequence of the number of grammatical relations they typically comprise; the simple ones are composed of only one grammatical function, but the potentially complex ones can be, and often are, made up of more.

In some or even most languages, it may be the case that layered bracketing is limited to a single level of embedding. In Mbyá Guaraní, however, two levels of embedding have been found in natural discourse. With Theme and Tail as adjunct functions occurring outside the pragmatic nucleus, we can get such additional configurations as outer Topic (inner Topic-Comment), as illustrated in 2 above, and Focus-(Topic-Comment) Tail.

Linguistic signals of pragmatic configurations are typically found among such phenomena as word order, intonation (especially breaks and peaks), occurrence of non-obligatory items (especially nominals), morphemic markers whose meaning indicates pragmatic functions or configurations, and cliticized elements serving positionally as boundary phenomena "in the cracks" between pragmatic functions. In general, a particular configuration will be recognizable to the extent that its component pragmatic functions: (i) satisfy prototypical notional features, such as definiteness for Topics, and (ii) appear as sharply delineated constituents of the sentence, by means of signals such as those listed above.

The following table summarizes the various types of pragmatic functions that are discussed in this paper:

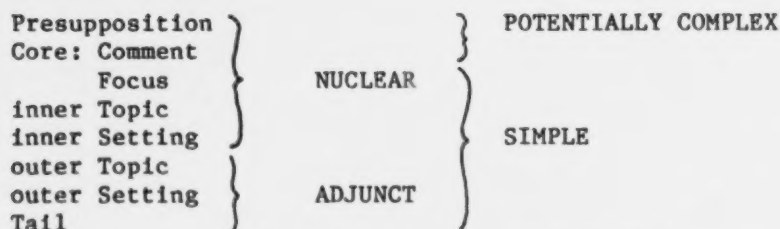


Figure 1: Pragmatic functions

In the framework being presented here, most if not all languages will have one of the nuclear pragmatic configurations as basic or unmarked. This basic configuration may have adjunct functions, such as (outer) Setting, outer Topic, or Tail, attached to it without taking on the status of a marked configuration. Marked configurations in a given language normally involve nuclear functions. These are the configurations which are used for special-purpose discourse functions. In addition, the Presentational configuration can be expected to be nonbasic (marked) in every language in which it occurs, because of its special discourse function.

Not every sentence will exhibit an overall pragmatic configuration. This is especially true for two classes of sentences: nondeclaratives and sentences in written or otherwise "planned" discourse. Nondeclaratives, such as the interrogatives **Is fat meat greasy?** and **Where is the nearest service station?**, commonly have special word orders which signal mood, thereby diminishing the coding devices available for pragmatic structuring. Planned discourse as well is characterized by, among other things, a heightened use of grammatical as opposed to pragmatic structuring. In these and similar cases, there may emerge no recognizable overall pragmatic construction for a given sentence. But in all languages, we can expect to find many sentences having an overall pragmatic configuration, with pragmatic functions occurring in a syntagmatic relationship. It is among these sentences that we search for a basic pragmatic configuration.

Pragmatic and grammatical structuring vie for the use of coding devices (principally word order, intonation, and morphological signals) that a language might have. There are at least four ways that a basic pragmatic configuration might relate to a basic ordering of grammatical relations in a given language:

- (1) It may be that a given language has a basic pragmatic configuration but no clear basic ordering of grammatical relations. An example of this is provided by the Brazilian Indian language Xavante (Ge family), which will be discussed in Sect. 5. In Xavante there is no clear basic ordering of grammatical relations, but there is strong evidence for Focus-Presupposition as the basic pragmatic configuration. In such a language,

pragmatic configurations form the principal means of structuring sentences, especially in regard to the order of constituents.

(ii) It may be that for a given language, a basic configuration of pragmatic functions coexists with a basic ordering of grammatical relations. The "languages in which word order functions pragmatically" discussed in Thompson 1978:20f (Russian, Mandarin, and Spanish) are all of this type. English, for example, may be described as both an SVO language and a Topic-Comment language, with S typically encoding Topic (Hockett 1958, 1966; Tomlin 1985). When a language has both of these types of basic configuration, they are necessarily mutually supportive of each other in some sense, due to limited coding resources. There are sometimes special constructions which help keep these two basic kinds of structuring simultaneously operative; English passivization has this function, since it involves a topicalized Patient becoming a preverbal S (Givón 1981:168ff).

(iii) A given language may have a basic configuration which is a composite of the two types of structuring, defined partly in terms of grammatical relations and partly in terms of pragmatic functions. This is exemplified by Nandi, a Nilo-Sharan language: "In Nandi, the unmarked position for the verb is sentence-initial, but the position of the nouns in a multi-participant sentence is determined by their pragmatic values", with the more rhematic (Comment or Focus) item preceding the more thematic one, in Prague School terminology (Thompson 1978:24f, citing Creider 1975).

(iv) A given language may have a basic ordering for grammatical relations, but no basic ordering for pragmatic functions. I do not know whether there are languages of this type, but the S-medial languages (VSO or OSV) are especially interesting in this regard. This point will be discussed further in Sect. 7.

We now turn our attention to three Brazilian Indian languages which illustrate three different basic pragmatic configurations: Mbyá Guaraní with Topic-Comment, Hixkaryana with Comment-Topic, and Xavante with Focus-Presupposition.

3 Topic-Comment in Mbyá Guaraní

Mbyá is a dialect of the Guaraní language, which in turn forms part of the Tupí family. Mbyá has approximately 2500 speakers in seven states of southern Brazil, and is spoken in northern Argentina and eastern Paraguay as well. It is distinct from the Avañeé dialect spoken by the general population of Paraguay.

Mbyá has subject prefixes and object prefixes. When the object is higher than the subject on the person hierarchy $1 > 2 > 3$, the subject prefix does not occur. Neither free subjects nor free objects are grammatically obligatory, and are absent from many sentences in natural discourse. There is no case marking for subject and object.

Mbyá has a basic SVO word order, with a high degree of word order freedom being utilized for purposes of discourse-pragmatic structuring. This structuring is described in Dooley (1982), from which the following citations and data are taken. In this language, "pragmatic structuring is realized primarily through a type of constituent structure" whose components are pragmatic functions; particles and pauses often occur "in the cracks" between such components to help delineate the configuration. Although most sentences have just one layer of pragmatic structuring, it is possible for the rightmost component to be itself structured into a binary configuration, and so on. In this way, three distinct levels have been found (307f).

These notions are illustrated in the following two examples which, in a narrative, were spoken by a young lady on two different occasions in order to indicate her choice of a husband; she, along with her sisters, was going to do this by throwing a flower at the one she wished. (The intonation center is indicated by capitals; an acute accent, when present, indicates a secondary peak of intonation; intonation breaks are indicated by a comma.)

- (7) Topic (-----Focus----- Presupposition)
 Xeé, peva'e ae re rIVE 'rã a-mombó.
 lSG that only at merely FUT lSG-throw²
 'As for me, merely at that one I will throw it.'

- (8) ----Focus-----
 Peva'e re rIVE 'rã ko
 that at merely FUT opinion

(Topic (Topic Comment))
 xeé yvotŷ a-mombó.
 lSG flower lSG-throw

'Merely at that one I will throw the flower.' (317)

In 7, the Comment component of the Topic-Comment configuration is in turn structured as Focus-Presupposition, and in 8 the Presupposition component itself has two layers of Topic-Comment structuring.

In Mbyá, the Focus-Presupposition configuration has been found in two common types of discourse-pragmatic conditions: (i) contrast (examples 7 and 8 above), and (ii) added-detail restatements, in which the Presupposition is restated from the preceding sentence (Dooley 1982:328f). Thus, it is a special-purpose configuration, occurring only under special discourse-pragmatic conditions.

Topic-Comment in Mbyá is much more common. In fact, given a prototypical association of Topic with grammatical subject in Mbyá and the basic word order SVO, Topic-Comment is the basic, unmarked pragmatic configuration (310):

- (9) Topic -Comment-
 Xe-ro o-kyty-PA.
 1SG-house 3-drip-completely
 'My house leaks completely.'

Under certain conditions there occur marked, especially informative Topics, but the order again is Topic-Comment. These occur (i) when there are new (sub)topics being introduced, (ii) when a pronominal expression is used to refer back to an entity which was just introduced, and (iii) in what Chafe (1976:35) refers to as double-focus contrast (his "focus" is not the same as the pragmatic function Focus). Marked Topics are recognizable by a sharply delineated Topic-Comment constituent structure, often with particles or pause "in the cracks" between the two components and a secondary peak of intonation on the Topic, as a counterpoint to the intonation center on the Comment. The first of the three types of marked Topic is illustrated by the glossed and labelled sentences in the following segment of procedural text:

- (10) a. 'From (material from) the woods I make a bow.

- b. I bring from the woods, wood which I will work.

- c. --Topic--
 I-para-rã,
 3-decoration-FUT

-----Comment-----
 a-jou guembe PI, takua reMBO.
 1SG-find guembe strip bamboo protrusion

For its decoration, I get strips of (the tendrils of the climbing plant) "guembe" and a type of fine bamboo with thorns.

- d. ----Topic---- -----Comment-----
 Guyrapa xã-rã, a-ke'o piNDO guive.
 bow string-FUT 1SG-cut:down palm also
 For the bowstring, I cut down a palm tree also.'

Based largely on data from English and Mbyá Guaraní, in Dooley 1982 I speculated on the possibility that Topic-Comment would turn out to be the unmarked pragmatic configuration in all languages. In the light of data about to be presented, that hypothesis no longer appears to be tenable.

4 Comment-Topic in Hixkaryana

Hixkaryana is a language of the Carib family with some 350 speakers (in 1979) who live about halfway between the Amazon and Guyana

(Derbyshire 1985:xiii). The verb is prefixed for agreement with subject and, when transitive, with object as well:

- (11) W-ama-no.
 1S:30-fell-IMM:PAST
 'I felled it (a tree)' (1985:3f).

(In certain of the Hixkaryana examples cited below, the gloss is for the word as a whole instead of morpheme-by-morpheme.)

Hixkaryana is the first carefully documented OVS language (see Derbyshire 1979a). In both Hixkaryana and other related Carib languages, Derbyshire (1981) attributes the change from an earlier SOV order to the grammaticalization of free subjects in the Tail position, where they supposedly occurred originally as "afterthought" elements. That is, elements would be described as being in the Tail position in the pre-grammaticalized phase; a major indicator of such grammaticalization is the phonological integration of the erstwhile Tail element into the preceding parts of the sentence, without an intonation break, as in 12 and 13.

- (12) Kuraha yonyhoryeno biryekomo.
 bow he:made:it boy
 'The boy made a bow.' (1985:31)

- (13) Kanawa y-aka-ye Tuhkoro.
 canoe 3S:30-make-DIST:PAST Tuhkoro
 'Tuhkoro made a canoe.' (1981:211)

In Hixkaryana, "the grammaticalization process has clearly reached a late stage, with OVS established as the basic order, but it may never be ... complete" (1985:103f), since, not infrequently, the sentence-final free subject is phonologically dislocated (1981:218). This is indicated by the comma in the following data:

- (14) Txemye hatf, Mawarye.
 he:poisoned:them HEARSAY Mawarye
 'Mawarye poisoned them.' (1965:67)

Right-dislocated components in Hixkaryana are not limited to subjects, but "are widely used in the language and with varying functions" (1985:104):

ADJUNCT:

- (15) Ekeh me wehxaha, atunano wya.
 sick:one DENOMLZR I:am fever by
 'I am sick with fever.' (1985:34)

NOMINAL OUTSIDE THE GRAMMATICAL NUCLEUS OF THE CLAUSE:

- (16) Koserhyehakon[†], romryen[†].
 I:was:afraid my:boyhood
 '(With reference to) my boyhood, I used to be afraid.'
 (1985:155)

DIRECT OBJECT:

- (17) Wenyhoryetxehkan ha, ~~tro~~ ha.
 I:finished:making:it INTENS that:thing INTENS
 'I have finished making that thing.' (1979b:78)

SECOND ITEM IN COORDINATION:

- (18) Hakrya wotxown[†] ha, koso heno komo.
 peccary they:shot:it INTENS deer group COLL
 'They shot peccary and some deer.' (1979b:78)

The Tail position is often used for what Derbyshire calls "frame of reference topics", defining these in the sense of Chafe's Mandarin Topics ("the frame within which the sentence holds"; see Sect. 2). This has already been illustrated in examples 14, 16, and 17; another example is given below (the parsing of Hixkaryana data into pragmatic functions is my own, RAD):

- (19) -----Comment----- ---Topic---
 Isna rma txko tyufa nkekon[†], oseryehr[†].
 to:there SAME DIMIN spitting she:did:it her:being:afraid

'(As an expression of) her being afraid, she was spitting into the little (pot).'

With this type of Topic, then, Hixkaryana has a Comment-Topic configuration. To show that Comment-Topic is actually the basic pragmatic configuration in the language, I will cite evidence for the following claims:

- (i) In the case of full NP subjects, there is an unmarked association of Topic with Subject.
 (ii) The order (O)VS is the most pragmatically neutral ordering of these elements when they are full NP's.

In this study I am disregarding constituent orderings with pronouns, since in Hixkaryana they have specialized discourse-pragmatic functions that are "totally different" from other referential devices in the language and not closely related to Topics (Derbyshire 1986:275ff). The question of clauses without full NP constituents is discussed in Sect. 6.

Derbyshire states that "The constituent that primarily carries unmarked theme is the subject, and this normally follows the verb (and object)" (1985:152). Here, "theme" is roughly equivalent to "Topic". The subject as unmarked Topic in Hixkaryana is demonstrated in a statistical study of "topic continuity" based on the model of Givón 1983. In this study, whereas the subject of a given clause has been referred to, on the average, 2.51 clauses earlier, the figure rises to 4.52 for direct objects and 7.49 for oblique objects (calculated from Tables 1, 6, and 10 of Derbyshire 1986). The evidence, then, is that subject is the grammatical relation in Hixkaryana which most consistently preserves continuity of reference. Since this is also one measure of topicality, this also indicates that subject in Hixkaryana has the highest degree of topicality among grammatical relations.

Derbyshire 1986 shows, among other things, that the (O)VS order is the most pragmatically neutral one in Hixkaryana when full (nonpronominal) NPs are present. The main points of the argument are summarized as follows. Continuous topics (those that occur medial and final in an episode) are most likely to be expressed either by verb agreement only or by VS order with an NP subject (267). In particular, "the postverbal noun phrase is a more continuous topic than the preverbal noun phrase" (275); "SV is a pragmatically marked order that is used sparingly even for the restricted functions which primarily characterize it" (279), namely, for "bringing into the register an entity that has not been on the scene for some time" (255). "The occurrences of VS", on the other hand, "range over almost every possible dimension of topic continuity and discontinuity" (279); this ordering is "used to express so many different discourse-pragmatic functions that it must be considered the most neutral of the coding devices" (272f). This kind of distribution for (O)VS is what one would expect from a basic or unmarked ordering. With the typical identification of S with Topic, this means that Comment-Topic is the basic or unmarked pragmatic configuration in Hixkaryana, when full NP's are in view.

It should also be mentioned that Hixkaryana uses Focus-Presupposition as a marked configuration. This involves an optional movement process which Derbyshire calls "fronting for emphasis" (1985:74). The following examples, taken from the page just cited, show this process applied to transitive (20) and intransitive subject (21):

- (20) --Focus--- -----Presupposition-----
 Okomkurusu b̄ryekomo heno yoskeko.
 bushmaster child dead it:bit:him
 'It was a bushmaster (snake) that bit the child.'

- (21) Focus- -----Presupposition-----
 Waraka haxa nehurkano asama yawo.
 Waraka CONTR he:fell trail on
 It was Waraka (not someone else) who fell on the
 trail.'

Uses of fronting for emphasis include contrastive emphasis, as in 20 and 21, and the introduction or reinstatement of an entity in discourse (1985:146ff), as already mentioned in the discussion of SV order. This configuration highlights items that the speaker wishes to present as being particularly informative. What follows the fronted constituent appears to have the pragmatic function Presupposition, containing information which the speaker feels that the hearer is not likely to challenge. Thus, Focus-Presupposition in Hixkaryana is a special-purpose configuration, occurring only under specific discourse-pragmatic conditions.

In summary, there are two identifiable pragmatic configurations for Hixkaryana sentences with free NPs: Comment-Topic and Focus-Presupposition. Topic-Comment does not appear to occur. Based on statistical and distributional evidence from texts, Comment-Topic seems clearly to be the basic or unmarked configuration. This fact correlates well with the basic ordering OVS for the language, given a typical association of S with Topic. The question of possible pragmatic configurations for sentences without NP arguments is discussed in Sect. 6.

5 Focus-Presupposition in Xavante

Xavante is a language of the Ge family with approximately two thousand speakers who live in eastern Mato Grosso near the headwaters of the Xingu and Araguaia rivers. In Xavante, "A predicate complex is the core of the clause. It is that part of the clause that begins with person-aspect proclitics and ends with the verb," together with possible modifiers postposed to the verb (Burgess 1986:28). The predicate complex, or simply the predicate, may include other elements, such as adverbial or direct object, between the person-aspect proclitics and the verb. These proclitics are coreferenced to the grammatical subject, which may or may not be represented as well by a noun phrase.

Basic word order is problematic in Xavante. Derbyshire (to appear) lists Xavante as ?OSV (SOV). Burgess states: "When both subject and object are identified by noun phrases, there is no overt distinction as to which is which either by affixation or by word order ... Their order relative to each other is determined by information or topical structure ... " (1986:28). The following examples are taken from McLeod & Mitchell 1977:125:

- (22) E waptsã ma tð adzö.
INTER dog 2/3 PERF he:beat
'Did he beat the dog?'
- (23) E ma tð waptsã adzö.
INTER 2/3 PERF dog he:beat
'Did he beat the dog?'

- (24) $\text{\textcircled{e}}$ ma t $\text{\textcircled{8}}$ adz $\text{\textcircled{8}}$ / wapts $\text{\textcircled{a}}$ h $\text{\textcircled{a}}$.
 INTER 2/3 PERF he:beat dog that:is
 'Did he beat it, that is, the dog?'

In 22, the object wapts $\text{\textcircled{a}}$ 'dog' occurs immediately after the interrogative marker $\text{\textcircled{e}}$ and before the predicate complex; in 23, it occurs within the predicate complex, immediately after the person-aspect proclitics ma t $\text{\textcircled{8}}$; in 24, it occurs in a Tail construction, together with the element h $\text{\textcircled{a}}$, which is discussed later in this section.

The Focus-Presupposition configuration occurs paragraph medial in Xavante. "The first sentence in a paragraph usually contains several items of new information. Further new information is added in the body of the paragraph, often one item per clause. The final sentence in a paragraph is often a summary one containing only given information" (Burgess 1986:33). The one item of new information is typically added as follows: "New primary information normally occurs before the predicate in a clause" (29f). (The word "primary" is used here in distinction to new secondary information, which is of lesser informativeness and hence is not formally distinguished from given information (30).) When new information is added sentence-initially, the remainder of the sentence takes on the pragmatic function of Presupposition.

To illustrate this, I give as example 25 the translation of an entire Xavante text (Text 1 from McLeod 1960). For reasons of space, I am omitting the text itself. I have parsed the recognizable Focus-Presupposition sentences into their pragmatic functions, and have rendered these in an English word order corresponding to the order of these constituents in Xavante.

- (25) a. Eunice and another are going to fly to Batovi
 (settlement).
 (New paragraph)
 b. Who with?
 c. ---Focus---
 With Alice.
 d. They are going to fly together.
 (New paragraph)
 (Question understood: Why are they going?)
 e. -----Focus----- -----Presupposition-----
 It is in order to see Helen that they will fly to
 -----Tail-----
 the dance, to the jaguar (festival).
 f. They will go over there.

- g. -----Focus-----
It is in order to see the jaguar (festival)

-----Presupposition-----
that they will fly there.

(New paragraph)
- h. Who will come here for them?
- i. Focus (Presupposition ellipsed)
Jim.
- j. Focus -----Presupposition-----
Jim will come here for them.
- k. ---Focus--- -----Presupposition-----
Next Monday he will come here for them.
- l. -----Focus----- Presupposition
Nharinha is the first one he will fly.
- m. They will fly together.

(New paragraph)
- n. Is Nharinha going to fly?
- o. She is flying to Cuiabá, and Nenê also.

(Background explanation)
- p. Dominga will stay here.
- q. -----Focus----- -Presupposition--
It's to her father that she will go.
- r. -----Focus----- -Presupposition--
It's to her father that she will go.
- s. So only Nharinha will fly (the first time).

(New paragraph)
- t. What (is Nharinha going) for?
- u. -----Focus-----
It's in order to get her eye scraped

---Presupposition--- -----Tail-----
that she is flying, since it hurts so much.
- v. -----Focus----- -----Presupposition-----
It's the painful part that he will scrape for her.

(New paragraph)

w. Does her eye hurt?

x. Yes, her eye hurts.

y. -----Focus----- --Presupposition--

It's because her eye hurts that she will fly,

-----Tail-----

so that someone can scrape it for her.

z. When Nharinha flies, the women will stay away for an indefinite period.

(New paragraph)

aa. Will the chief stay by himself then?

bb. ---Focus--- Presupposition -----Tail-----

By himself he will stay, the man alone,

-----Tail-----

the man alone.

cc. Kosisababa will go away to the fields.

dd. She will stay there.

ee. When she (Nharinha) returns, she will come back here.

ff. Her husband, as soon as he finishes the house-building, he will go there.

gg. As soon as he finishes the house-building, he will go there, Tsiriwaruwẽ (husband's name).

The paragraph-initial questions (b, h, n, t, w, aa) are from the narrator to himself. This is a common paragraph-initial device in Xavante, along with sentences giving several new items of information (Burgess 1986). Sentences d, f, m, and s are summary statements; as per Burgess, these are typically paragraph final. Tail elements, which generally give additional information, occur in e, u, y, and bb. In at least two of the Focus-Presupposition configurations, j and r, the Focus does not consist of new information at all, but is repeated from the preceding sentence. Not only the content, but also the Focus-Presupposition configuration as well is repeated from the preceding sentence. Burgess suggests that this may signal the conclusion of a high-level topic (Burgess 1986:34). Sentences bb through gg appear to be part of some type of coda to the text which provides related explanation.

Speaking from a non-Xavante viewpoint, the most prominent feature of this text is the high frequency of Focus-Presupposition configurations occurring paragraph-medially. The first impression of an English speaker is that these sentences overplay the introduction of new information. This is because, in English, this configuration is used only in highly specialized contexts, when the Focus information rates rather drastic highlighting:

- (26) ---Focus--- --Presupposition--
It was Fred that married Laura (not Max).

This configuration is common in Xavante because it has taken on the role of the basic or unmarked pragmatic configuration. Cross-linguistically, it is paragraph-medial sentences which best display neutral or unmarked patterns; paragraph-initial and paragraph-final sentences tend to be special-purpose types (Givón 1983).

The actual Xavante data for paragraph h through m are given below as example 27.

- (27) h. È wa dza wê tsô ʔwitsi?
INTER who FUT here them:for come
'Who will come here for them?
- i. --Focus-- (Presupposition ellipsed)
Dzemi-hã.
Jim-that:is
Jim.
- j. ---Focus--- -----Presupposition-----
Dzemi-hã tê dza wê tsô ʔwitsi.
Jim-that:is 3 FUT here them:for come
Jim will come here for them.
- k. -----Focus-----
romhuriduridzep-amo-na
Monday-other-on
- Presupposition-----
tê dza wê tsô ʔwitsi.
3 FUT here them:for come
- Next Monday he will come here for them.
- l. -----Focus----- Presupposition
Nharinha-dza'utsi tê dza atsamrô.
Nharinha-first 3 FUT fly
Nharinha is the first one he will fly.
- m. Tê dza atsamrô-dzahure.
3 FUT fly-DUAL
They will fly together.'

Although Focus-Presupposition is the unmarked pragmatic configuration in Xavante, it is not the only configuration. In Sect. 3 of her paper, Burgess discusses Topic-Comment. This configuration is quite noticeable with interrogatives which begin a discourse or paragraph (1986:36):

- (28) ----Topic----- -----Core-----
 Pi'o-nori-hã, @ momo tẽ ai'aba're?
 woman-PL-that:is INTER where they go
 'The women, where are they going?'

- (29) -Topic- -----Core-----
 Donaudu, @ mame tẽ ñamra?
 Donald INTER where he live
 'Donald, where does he live?'

(In 28 and 29, I have used the term "Core" instead of "Comment" because it is less suggestive of the declarative mood.)

Topic-Comment (Topic-Core) configurations also occur in descriptive sentences involving a type of rhetorical question (loc.cit.):

- (30) -----Topic----- -----Core-----
 Udzõ-hã barana-hã, @ I-ro'o-baihõirê?
 light-that:is night-that:is INTER which-burn-many
 'Lights at night, wow! are there ever a lot!'

The Topics in 28-30 are analyzed by Burgess as "marked Topics", whose function is to introduce or reintroduce entities into the discourse. Since they are outside of the syntactic nucleus of the clause, they are analyzed as outer Topics in the framework of this paper. One of the linguistic signals of their occurrence is phonological left dislocation, with an intonation break before the Core. Another is the occurrence of the enclitic element *-hã*. In McLeod (1974), this morpheme is variously glossed as 'it is', 'specifier', and 'focuser'. In the examples presented in this paper, it occurs following Tail (24), Focus (27i, j) and outer Topic (28-30). That is, it seems to occur with pragmatic functions giving new or, in some sense, highly informative content. Further study is needed to determine whether this is an adequate characterization.

According to Burgess, there is also an unmarked (inner) Topic-Comment configuration in Xavante: "In a clause in which all information is new, the first element is the topic. If the clause is intransitive, the subject is most likely to be the topic. If the clause is transitive, either subject or object may be topic, and their order relative to each other is determined by which is the topic, the topic being the first element" (1986:38). This kind of configuration is illustrated in examples 31 and 32; the Topics are, respectively, subject and object:

- (31) Topic ---Comment--
 Litsi ma tē-dzada.
 Lici she leg-burn
 'Lici burned her leg.'
- (32) Topic -----Comment-----
 Toptō wahi matē ti-tsa.
 Toptō snake it her-bite
 'A snake bit Toptō.' (38)

Burgess also identifies unmarked Topics in sentences with given information (1986:39), but it is not clear that there is any linguistic evidence of their being distinct pragmatic functions. Recognizable Topic-Comment configurations do occur in Xavante, but only when Topic is new information. On this basis, we classify them as marked, special-purpose configurations. It is not known whether there are formal distinctions in Xavante between Topic-Comment and Focus-Presupposition, but the nonpresuppositional character of the Comment component seems to distinguish between the two configurations.

In Xavante, then, although it is problematic to establish a basic order among the grammatical relations S, V, and O, available descriptions and text data point rather clearly to Focus-Presupposition as the basic pragmatic configuration.

6 "V-only" languages

"V-only" is a term used by Payne (1986) to describe a situation in a language in which there is a statistical predominance of clauses having no free S or O elements. Such languages have presented problems in determining the basic order of grammatical relations (Brody 1982). One such language is Yagua of Peru (unitary member of the Peba-Yaguan family) (Payne 1986), and another is Hixkaryana of Brazil (Carib) (Derbyshire 1986). Both are V-only in regard to the most neutral sentence type in natural discourse, but both have identifiable basic orderings of grammatical relations when free S and O nominals do occur: Yagua as VSO (Payne, 460) and Hixkaryana as OVS (Derbyshire 1986:281). For languages of this type, basic word order can be investigated, as for other languages, in those sentences where both S and O occur as free nominals, and thus, even though many sentences are V-only, "we can maintain a basic word order concept that is valid for all known languages" (Derbyshire, to appear).

For such languages, there are methodological problems in determining a basic pragmatic configuration. There appear to be three possibilities:

- (1) The language could be analyzed as having only one pragmatic function in the basic configuration, and that would be Focus/Comment (recall that I am treating Focus and Comment as varieties of the same basic pragmatic function, Sect. 2).

Thus, the basic pragmatic configuration for all V-only languages would be Focus/Comment-only.

- (ii) The same procedure could be followed for determining the basic pragmatic configuration as for determining basic word order; that is, the V-only sentences would be ignored, and the basic pragmatic configuration for the (rest of the) language would be the least marked one when there are two or more pragmatic functions in evidence.
- (iii) Since there are obvious acts of reference performed in most V-only sentences, the possibility could be explored that these sentences have an ellipsed pragmatic function. That is, if there is a least marked pragmatic configuration, say Topic-Comment, for sentences in the language with free NP arguments, then (Topic)-Comment could be claimed as the basic configuration for the language as a whole.

Alternatives (i) and (ii) are less than ideal, in the sense that each fails to take into account a certain class of sentences in the language; (i) ignores sentences with free NPs, and (ii) ignores the V-only sentences. Alternative (iii) is only viable on the condition that there is a least marked pragmatic configuration among those sentences with free NPs, but given that the condition holds, it is the best option of the three.

In Hixkaryana, for example, we have shown in Sect. 4 that Comment-Topic is the least marked configuration for sentences with free NPs. The basic or unmarked pragmatic configuration for the language, then, would be Comment-(Topic).

7 Toward universals of pragmatic structuring

If it is true that "discourse universals tend to be 'more universal' than syntactic structure" (Givón 1984:129), then one is certainly justified in looking for universals in regard to pragmatic configurations. In Sect. 1, it was mentioned that early treatments of this area sometimes cited characteristics of human cognition which suggested the universality of something like the following:

- (1) "The ordering Topic-Comment is the only basic pragmatic configuration in human languages; other configurations only occur as marked choices."

This is a hypothesis I favored in Dooley 1982.

More recently, Givón has posited a universal in the opposite direction. Not only does he state that "topic-comment is not the unmarked universal word-order", but he asserts that

- (33) Topic-Comment "is the marked universal word-order, the one used

when the topic is discontinuous, surprising, less obvious, less predictable" (Givón 1984:128).

He then goes on to put forward characteristics of human cognition in favor of his own position: "But this only makes sense. If one has difficulty establishing a new topic, or if one suspects that the hearer is likely to experience such a difficulty, the most sensible strategy is first to make sure that the topic is firmly established, and only then to come up with the new information" (loc. cit.)

Evidence has been put forward in this paper to show that neither 1 nor 33 has universal validity. Topic-Comment, Comment-Topic, and Focus-Presupposition all serve as basic pragmatic configurations in human language. The following table summarizes the major pragmatic configurations discussed in this paper:

	Basic word order	Basic pragmatic configuration	Marked pragmatic configuration
Mbyá	SVO	Topic-Comment	Focus-Presupposition
Hixkaryana	OVS	Comment-Topic	Focus-Presupposition
Xavante	?OSV	Focus-Presupposition	Topic-Comment

Figure 2. Basic and marked orderings

Since neither 1 nor 33 hold, universals of pragmatic configuration must be formulated along other lines. The following three statements are offered as possibilities:

- (34) In all languages, sentences can be found having an overall pragmatic configuration in which there are pragmatic functions in a syntagmatic relationship.
- (35) All languages without a basic ordering of grammatical relations (S, O, V) have a basic configuration of pragmatic functions.
- (36) All languages have a basic configuration of pragmatic functions.

Note that 36 is stronger than either 34 or 35 in that it implies both of them, but neither 34 nor 35 implies the other.

Statement 34 would be a trivial consequence of the universal occurrence of any particular type of pragmatic configuration. For example, if either of the following statements proved true:

- (37) a. Every language has either a Topic-Comment or a
Comment-Topic configuration;
b. Every language has a Focus-Presupposition
configuration,

then 34 would follow as a corollary. Both 37a and 37b hold in the three Brazilian languages discussed in this paper.

Statement 35 relates especially to the question of "free word order languages". Concerning such languages, Thompson states: "I am suspicious of the notion 'free word order'. A language which is described this way will generally turn out, upon closer investigation, to be one in which pragmatic factors determine the position of major constituents; the task of the linguist working on such a language is to specify what these factors are and how they interact" (1978:23). Given the powerful coding tool that an unmarked ordering of sentence constituents is, it would be *a priori* unlikely that a language would fail to take advantage of it either in pragmatics or in grammar.

For statement 36, certain V-only languages (Sect. 6) may present an obstacle. If, in a certain language, the major part of the sentences in natural discourse are of the V-only type, then it is possible that among the remaining sentences there may be no configuration which is least marked discourse-pragmatically. In Yagua, when pragmatic criteria are taken into account, there is a least marked word order for sentences with free NPs, and it is VSO (Payne 1986:453). However, it is not known whether there is any particular configuration of pragmatic functions which corresponds to this.

Due to topic properties often associated with grammatical subject, it may be that

- (38) Languages with medial S in their basic word order
(VSO and OSV languages) do not have either Topic-Comment or
Comment-Topic as a basic pragmatic configuration.

Xavante (?OSV) has Focus-Presupposition as its basic configuration and Topic-Comment as a marked configuration. As just mentioned, the case for Yagua (VSO) is unclear.

8 Conclusion

Evidence presented in this paper indicates that neither Topic-Comment nor the opposite order of Comment-Topic can be universally taken as a basic/unmarked pragmatic configuration in human languages. Rather, it appears that languages adopt a basic configuration from a certain small set of possibilities, all of which have plausible explanations in terms of human cognition and discourse-pragmatics. These include Topic-Comment, Comment-Topic, and Focus-Presupposition. It would be a research program of obvious value to determine what statements regarding pragmatic configurations actually do have universal validity. It would also be important to investigate why one or another pragmatic

configuration "catches on" in a given language at a given time and becomes the basic configuration, while others function as marked, special-purpose configurations, or else do not occur at all.

The basic question for the investigation of pragmatic configurations in language is the same as what Perlmutter gives as the basic question of linguistics as a whole: "In what ways do natural languages differ, and in what ways are they all alike?" (Perlmutter 1980:195). And, especially for a part of language like pragmatics that interfaces in an obvious way with extralinguistic phenomena, the further question needs to be asked as well: "Why do these differences and commonalities occur where they do?"

Notes:

1. Thompson, after quoting from certain Prague School linguists, states that "In English, the positions of those sentence elements is [sic] at the disposal of grammatical determinants virtually to the exclusion of pragmatic factors" (1978:25). Since English gave major if not primary impetus to early recognition of the Topic-Comment configuration (Sapir 1921, Hockett 1958:191, 201f) and further studies such as Tomlin 1985 appear to bear out the influence of discourse-pragmatic factors on subject assignment and hence on initial position in many sentences, Thompson's assessment seems too strong.
2. The following abbreviations are used in the glosses of data in this paper:

COLL	collective	NEG	negative
CONTR	contrast	O	object
DENOMLZR	denominalizer	PERF	perfective
DIMIN	diminutive	PL	plural
DIST:PAST	distant past	PROG	progressive
EXCL	exclusive	S	subject
FUT	future	SG	singular
HABIT	habitual	STAT	stative
IMM:PAST	immediate past	1	first person
INTENS	intensifier	2	second person
INTER	interrogative	3	third person

References:

- Andrews, Avery. 1985. The major functions of the noun phrase. *Language typology and syntactic description*, vol. 1: Clause structure, ed. by Timothy Shopen, 62-154. Cambridge: Cambridge U. Press.
- Brody, Jill. 1982. Some problems with the concept of basic word order. Paper presented at the Fifty-seventh Annual Meeting of the LSA, San Diego, CA.

- Burgess, Eunice. 1986. Focus and topic in Xavante. Sentence initial devices, ed. by Joseph E. Grimes, 27-41. Dallas: SIL and Arlington: U. of Texas at Arlington.
- Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. Subject and Topic, ed. by Charles N. Li, 25-56. New York: Academic Press.
- Chomsky, Noam. 1971. Deep structure, surface structure, and semantic interpretation. Semantics, ed. by Danny D. Steinberg and Leon A. Jakobovits, 183-216. Cambridge: Cambridge U. Press.
- Comrie, Bernard. 1981. Language universals and linguistic typology. Oxford: Blackwells and Chicago: U. of Chicago Press.
- Creider, Chet. 1975. Thematization and word order. Paper presented at the LSA Annual Meeting.
- Daneš, František. 1967. Order of elements and sentence intonation. To honor Roman Jakobson: Essays on the occasion of his seventieth birthday, 499-512. The Hague: Mouton. [Reprinted in Bolinger 1972:216-32, from which my citations are taken.]
- Derbyshire, Desmond. 1965. Textos Hixkaryãna. Museu Paraense Emílio Goeldi, Publicações Avulsas No. 3. Belém, Pará.
- _____. 1979a. Hixkaryana syntax. Ph. D. dissertation. University of London.
- _____. 1979b. Hixkaryana. *Lingua Descriptive Series* 1. Amsterdam: North-Holland.
- _____. 1981. A diachronic explanation for the origin of OVS in some Carib languages. *Journal of Linguistics* 17.209-20.
- _____. 1985. Hixkaryana and linguistic typology. *SIL Publications in Linguistics* 76. Dallas: SIL and Arlington: U. of Texas at Arlington.
- _____. 1986. Topic continuity and OVS order in Hixkaryana. *Native South American discourse*, ed. by Joel Sherzer and Greg Urban, 237-305. Berlin: Mouton de Gruyter.
- _____. to appear. Morphosyntactic areal characteristics of Amazonian languages. *IJAL*.
- Dik, Simon C. 1978. *Functional grammar*. Amsterdam: North-Holland.
- Dooley, Robert A. 1982. Options in the pragmatic structuring of Guaraní sentences. *Lg.* 58.307-31.

- Givón, Talmy. 1979. On understanding grammar. New York: Academic Press.
- . 1981. Typology and functional domains. *Studies in Language* 5.163-93.
- . 1983. Topic continuity in discourse: an introduction. Topic continuity in discourse: a quantitative cross-language study, ed. by T. Givón, 1-41. Amsterdam: John Benjamins.
- . 1984. Universals of discourse structure and second language acquisition. Language universals and second language acquisition, ed. by William E. Rutherford, 110-35. Amsterdam: John Benjamins.
- Hockett, Charles F. 1958. A course in modern linguistics. New York: MacMillan.
- . 1966. The problem of universals in language. Universals of language. 2nd edn., ed. by Joseph H. Greenberg, 1-29. Cambridge, MA: MIT Press.
- Keenan, Edward L. III. 1978. The syntax of subject-final languages. Syntactic typology: studies in the phenomenology of language, ed. by Winfred P. Lehmann, 267-327. Austin: U. of Texas Press.
- Li, Charles N. and Sandra A. Thompson. 1976. Subject and topic: a new typology of language. Subject and topic, ed. by Charles N. Li., 457-89. New York: Academic Press.
- Mallinson, Graham and Barry J. Blake. 1981. Language typology: cross-linguistic studies in syntax. Amsterdam: North-Holland.
- McLeod, Ruth, ed. 1960. Two Xavante texts. *Arquivo Lingüístico*, number 98. Brasília: Summer Institute of Linguistics.
- . 1974. Paragraph, aspect, and participant in Xavante. *Linguistics* 132.51-74.
- and Valerie Mitchell. 1977. Aspectos da língua Xavante. Brasília: SIL.
- Payne, Doris L. 1986. Basic constituent order in Yagua clauses: implications for word order universals. *Handbook of Amazonian languages* 1, 440-65. Berlin: Mouton.
- Perlmutter, David M. 1980. Relational grammar. Syntax and semantics, vol. 12: current approaches to syntax, ed. by Edith A. Moravcsik and Jessica R. Wirth, pp. 195-229. New York: Academic Press.
- Reinhart, Tanya. 1982. Pragmatics and linguistics: an analysis of sentence topics. Bloomington, IN: Indiana U. Linguistics Club.

- Sapir, Edward. 1921. Language: an introduction to the study of speech. New York: Harcourt, Brace & World.
- Thompson, Sandra A. 1978. Modern English from a typological point of view: some implications of the function of word [sic] order. *Linguistische Berichte* 54.19-35
- Tomlin, Russell S. 1985. Interaction of subject, theme, and agent. *Beyond the sentence: discourse and sentential form*, ed. by Jessica R. Wirth, 59-81. Ann Arbor, MI: Karoma.
- Watters, John R. 1979. Focus in Aghem: a study of its formal correlates and typology. *Aghem grammatical structure. SCOPIL 7*, ed. by Larry M. Hyman. Los Angeles: UCLA, pp. 137-97.

THE SYLLABLE STRUCTURE OF SERI*

Stephen A. Marlett

- 1 Introduction
- 2 Consonants and vowels
- 3 Onsets
 - 3.1 Syllables without onsets
 - 3.2 Simple onsets
 - 3.3 Complex onsets
 - 3.4 Round consonants
- 4 Nuclei
 - 4.1 Long vowels
 - 4.2 Stress and a restriction on complex nuclei
 - 4.3 A restriction on vowel clusters
- 5 Codas
 - 5.1 Simple codas
 - 5.2 Complex codas
 - 5.3 Round consonants in codas
- 6 Word-medial consonant clusters
- 7 Syllable overview: Uninflected words
- 8 Derived words
 - 8.1 New clusters
 - 8.2 Resyllabification
 - 8.3 o-Insertion
 - 8.4 V-Insertion

1 Introduction

The syllable structure of Seri is the focus of study of Moser and Moser 1965, and is also briefly discussed in Marlett 1981 and Marlett and Stemberger 1983. None of these works, however, adequately describes the important features of the Seri syllable. In this paper I begin to rectify this situation by presenting a more detailed discussion of this aspect of the Seri language. While various theoretical issues are discussed, my main objective is to provide a comprehensive understanding of Seri syllable structure and facts which are relevant to it. In some cases I argue that certain data support a particular structure over another, but in some cases alternative treatments may also be compatible with the facts. These facts are relevant to a number of issues, including the sonority hierarchy, the internal structure of syllables (especially the nucleus), the representation of long vowels, the representation of round consonants, the CV tier, the relationship of stress to syllabification, and the distinction between the syllable structure of inflected words and uninflected words.

In the first several sections below I discuss primarily patterns which can be found in monomorphemic words. The syllable structure of

such words, while more complex than many other languages, is still relatively restricted. Sect. 7 summarizes these facts. In Sect. 8 I take up inflected words. Unlike many other languages where inflected words generally have the same syllable structure as uninflected words, in Seri these derived words have radically different syllables. Affixation is responsible for all four-segment onsets and codas, and for almost all three-segment onsets described in previous studies. Various constraints and restrictions which are developed for uninflected words turn out to be irrelevant.

2 Consonants and vowels

The segmental inventory of Seri is given in (1).

(1) Consonants	Stops	p	t			k	k ^w		
	Fricatives	f	s	ʃ	ʂ	x	W	X	X ^w
	Nasals	m	n						
	Glides					y			?
Vowels		i	e	a	o				

In addition to the consonants listed here there is a voiced lateral, which is almost nonexistent in the present-day language,¹ and a flap r, which is limited to loanwords.² I do not deal with these in the following study. The symbol f represents a bilabial voiceless fricative, ʃ a voiceless retroflexed alveopalatal fricative, W a voiceless spirantized w, and X a voiceless uvular fricative.³ The round consonants k^w, W, and X^w have an extremely limited distribution due to some fairly transparent historical developments. They are therefore basically ignored in what follows, except in Sect. 3.4 and 5.3, in which their distribution is described.⁴ Glottal stop is listed as a glide since it patterns with the sonorants, as shown below.

With respect to the vowels, the symbol e represents a low front vowel [æ]; its frequency in words is much lower than the other vowels. The language allows a vocalic segment to occupy two vowel positions: i.e., Seri has long vowels. This analysis, which is different from that of Moser and Moser 1965, is discussed in Sect. 4.1 below.

Stress generally occurs on the first syllable of the root and is discussed in Sect. 4.2. However, I write it below whenever it is not on the first syllable of the word.

3 Onsets

3.1 Syllables without onsets

Although syllables without onsets are much rarer than those with onsets in Seri, they do exist.

In (2) I present words which begin with a vowel, divided according to the stress pattern and vowel length. Long vowels do not occur in unstressed syllables (see Sect. 4.2 below for details regarding stress and this constraint); therefore there is no sixth pattern V:CV. The parenthesized elements indicate that words of this type are found only with a morpheme boundary following the vowel.

(2) Monosyllabic words

VC...

ak 'the'

ix 'different'

oX 'thus'

V:C...

a:Xt 'plant (sp.)'

e:x 'strength'

i:kx 'coarse sand'

o:t 'organpipe cactus (sp.)'

Bisyllabic words

VC...

ašox 'star'

(i-tak)⁵ 'his/her bone'

V:C...

a:mo 'far'

e:nim⁶ 'knife'

(i:-stoX) 'their breath'

o:šXtim 'itching rash'

VCV...

(a-sā:k) 'his son'

(i-tit) 'his/her head'

otāk 'toad'

A word such as **koai** (NOM-D-delouse) might at first glance appear to be an example of a word containing a non-initial syllable without an onset. Nevertheless, in the analysis presented below, as in Moser and Moser 1965, syllables without onsets occur only word-initially. This word is analyzed as involving a single syllable with three vowel positions. Evidence for this analysis is presented in Sect. 4.3.

3.2 Simple onsets

Any consonant (excluding the round ones from discussion, as indicated in Sect. 2) may occur in a simple (i.e. single segment) onset. The fricatives **f**, **ɬ**, and **x** are very rare as word-initial onsets, but they occur with greater frequency word-medially. (I make note of the skewed distributions that I have observed, but my analysis does not necessarily deal with all of them.) Word-medial consonant clusters are taken up in Sect. 6. (The glosses for the words in (3) are included in the appendix.)

(3)	CV...	CV...	...VCV	...VCV
p	pak	paɪt	Xapɔ:	?epem
t	tis	taxɪst	potɔ:	iti
k	ke:ɪX	kotɔpis	?akɪmet	ste:kox
f	fe:ɪ	---	koffkx	i:fa
s	set	sanɔ:kx	kasɔmikt	?aso
ɪ	---	ɪana:kx	paɪt	Xe:te
ʃ	ʃa:p	ʃakɔ:m	aʃɔ:k	aʃox
x	xoene	---	paxɔ:s	ko:xox
X	Xepe	Xaʃɔ:p	moXɪma	iXox
m	mo:snɪ	moXɪma	tamɔix	?eme
n	nop	naxɔ:	konɔ:	no:ni
?	?ast	?amɔ:n	?aʃɔ:t	kaʃa
y	ye:n	yatɔ:m	koyɔko	Xpeyo

3.3 Complex onsets

Complex onsets exist in Seri, although they are very much less common than simple onsets. (Remember that we are dealing with monomorphemic words here.) The attested word-initial onsets are given in (4), but most of them occur in only a few monomorphemic words (often less than three). Complex onsets which begin with **X** are a bit more common, especially **Xp**, **Xt**, and **Xn**. (As we will see later, many other clusters are commonly found when affixation is involved.) The cluster **sk** occurs in the prefix **ska-**, but not elsewhere. Complex onsets always consist of non-identical consonants. **f** and **?** do not occur in complex onsets within a single morpheme (with the exception of the first person singular intransitive prefix **?p-**, which is discussed in Sect. 8.2).

The examples in the columns in (4) all display the complex onsets of stressed syllables. The same onsets in unstressed word-initial syllables are much rarer, understandably so since words of this stress pattern are also much rarer; examples are given below each table. (Glosses are given in the appendix.) (Word-medial consonant clusters are discussed in Sect. 6.) As these tables imply, no complex onset has a glide or a nasal as the first segment. This fact is important and I return to it later.

Word-initial three-segment onsets occur in only two monomorphemic words: **ptkɔmn** 'lobster' and **Xpkɪmɪ** (nickname).

(4) Word-initial complex onsets

a. Stop as second segment

	-p	-t	-k
p-	*	pta:kt	
t-		*	
k-	kpo:t	ktam	*
s-	spitx	stak	ska-
š-			škaXt
ɬ-			
x-			xkoa
X-	Xpašx	Xti:p	Xkokni

In unstressed syllables: ktamó:, Xpaká:fk, Xtamá:ixa

b. Fricative as second segment

	-s	-š	-ɬ	-x	-X
p-	psa:k				pXa:ʔom
t-		tšɬiʔ		txamoxa	
k-	ksaɪ				
s-	*		sɬe:		
š-		*			
ɬ-			*		ɬXe:kox
x-				*	
X-			Xɬoɬo	Xxi:	*

In unstressed syllables: tšikáktox

c. Sonorant as second segment

	-m	-n	-y
p-		pna:kox	pyoke
t-	tm-		
k-	kma:m	knaxox	kyaXa
s-		sna:šX	
š-		šnapX ^w ɬ	
ɬ-			
x-			
X-		Xnois	

In unstressed syllables: kmaxí:k, Xnaʔá:iɬ

It should be noted that there are no complex onsets such as **pp** and **tt** cited above. A Seri onset cannot consist of identical elements morpheme-internally. And just as a single consonantal segment in Seri

cannot be associated with two consonant positions in the onset, it cannot be associated with two consonant positions in rhymes, or even intra-syllabically. In other words, there are no geminate consonants. (But see Sect. 8.1.)

Some facts with respect to complex onsets in Seri are not predicted by current theories of the syllable which explicitly incorporate a universal sonority scale. A fairly detailed scale is proposed in Selkirk 1984, for example, but I will assume here a version which refers only to major classes: Obstruents < Nasals < Liquids < Glides < Vowels. Harris (1983) appeals to such a scale to rule out certain disallowed onsets in Spanish, on the assumption that the syllable typically increases in sonority as one moves from the edge toward the nucleus. An onset consisting of two obstruents is considered marked by this scale, since "in the unmarked case the left-to-right order of [+consonantal] segments in onsets is obstruent-nasal-liquid" (p. 21). The onset obstruent-obstruent does not occur in Spanish. It does occur in Seri, however, as we have already seen.

The universal sonority scale correctly predicts that nasal-nasal and nasal-obstruent onsets should not occur in Seri. I extend this prediction to naturally exclude all sonorant-obstruent and sonorant-sonorant onsets. Therefore the only fact about Seri syllable onsets that is marked, and which therefore needs to be explicitly allowed for, is the existence of obstruent-obstruent onsets.⁸

3.4 Round consonants

Round consonants occur only rarely in word-initial onsets of monomorphemic words. This is understandable since round consonants apparently arose historically from the loss of an *o* (see Sect. 5.3), and this loss apparently happened in a very restricted fashion pretonically. The only clear examples are the following: *k^wset* 'plant (sp.)', *k^wtep* 'plant (sp.)', *k^wsañi* 'plant (sp.)', *k^wleɬ* 'parrot', *k^wsek^wta* (place name).

4 Nuclei

As stated in Sect. 2, Seri has four vowels which occur both 'long' and 'short'. The vowel *e*, the low front unrounded vowel, is less common than the others. Long vowels, and also clusters of identical vowels, are simply phonetically longer in duration than short vowels. Restrictions on long vowels are discussed below.

4.1 Long vowels

Long vowels in Seri must be represented generally as in (5a) rather than as in (5b). That is, long vowels are a single vowel on the segmental tier but occupy two vowel slots on the CV tier. We know this from the way in which such segments interact with the phonology and morphology, as I show presently.⁹



The evidence for (5a) is considerable. For example, there is a general rule that deletes a root-initial short low vowel when it follows another vowel. This is shown by the forms in (6).

- (6) k + aitom --> kaitom (NOM-speak)
 t + aitom --> taitom (RL-speak)
 po + aitom --> poitom (IRR-speak)
 si + aitom --> siitom (IRR-speak)

This deletion rule does not apply to long low vowels, as the forms in (7) show.

- (7) k + a:ix --> ka:ix (NOM-undulate)
 t + a:ix --> ta:ix (RL-undulate)
 po + a:ix --> pa:ix (IRR-undulate)
 si + a:ix --> sa:ix (IRR-undulate)

Thus the rule deleting a low vowel must refer to 'short' vowels only. If long vowels were generally represented as in (5b), the statement of the deletion rule would be complicated considerably. This same complication would also have to be repeated elsewhere. Other facts which establish this point are discussed in Marlett 1981 and Marlett and Stemberger 1983.

Generally only vowels may be associated with V positions in Seri. Syllabic nasals occur in a very restricted environment. These are discussed in Sect. 8.4.

4.2 Stress and a restriction on complex nuclei

The nature of primary stress placement in Seri is relevant to ensuing discussion. On the word level, stress generally occurs on the leftmost syllable of the root. The presence of various kinds of affixes does not affect the placement of stress. Some verbs and many nouns, however, display stress patterns which are at odds with this generalization. In the case of verbs, this is generally due to the lexicalization of a prefix-root combination. In the case of nouns, this is generally due to the lexicalization of a noun-noun compound. Nevertheless, the following observation is true in all cases:

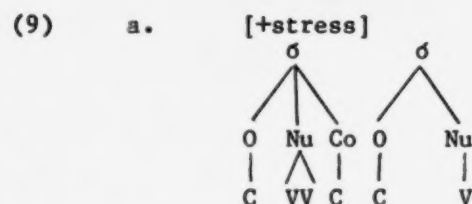
- (8) An unstressed syllable cannot contain a complex nucleus.

(The matter of the syllabification of vowel sequences is discussed below; (8) is correct only if my analysis of such is adopted.)

As far as word-level phonology is concerned, a complex nucleus is found only in the primary stressed syllable of the word. Or, to state it in reverse, stress is never assigned to a syllable with a simple nucleus if the root contains a syllable with a complex nucleus.

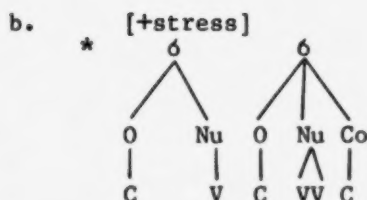
Stress is therefore assigned to the leftmost syllable of the root unless another syllable in the root contains a complex nucleus (or if stress is lexically marked). In nouns such as **Xpeyo** 'sailfish' and **sle:kox** 'heron (sp.)' stress is assigned to the leftmost syllable by the general rule. In nouns such as **Xapδ:** 'sea lion' and **paxā:s** 'squid' stress is assigned to the second syllable because stress could not be assigned to the first since another syllable has a complex nucleus. In nouns such as **moXina** 'yesterday' and **kotōpis** 'shellfish (sp.)' stress must be simply assigned by counting the first syllable as extrametrical, or by marking it in the lexicon.¹⁰

Given the stress assignment rule and/or generalization (8), we expect to find structures such as those in (9a) and not those in (9b).



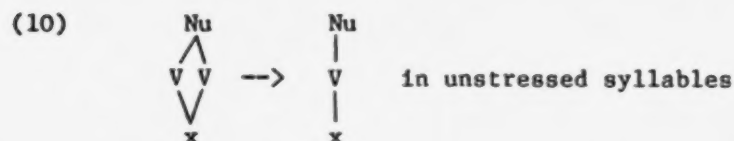
Examples:

kōanʔa 'tarantula'
nā:pXa 'buzzard (sp.)'



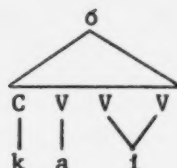
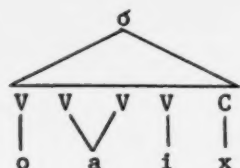
* tōpaix
* kīta:s

But (8) is not simply a static generalization. A rule simplifying certain complex nuclei in the absence of stress is also needed, both at the word level and the phrase level. Some prefixes arguably have long vowels in underlying form, but do not have long vowels in phonetic form. In compounds, only the rightmost member retains primary stress, and a long vowel in the other member must shorten. In unstressed clauses, long vowels shorten. Therefore a rule such as (10) is needed: Two V slots on the CV tier are collapsed into one if they are associated with a single vowel (represented by x) and occur in an unstressed syllable.

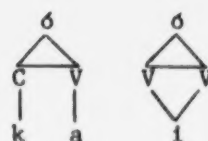
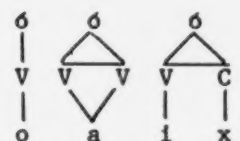


Generalization (8), rule (10), and the stress rule lead me to take the position here that roots such as *-o:ix* 'hang free' and words such as *kai:* 'mature' are monosyllabic structurally (and also phonetically in most cases), as shown in (11a), rather than being polysyllabic, as in (11b-d). (See Clements and Keyser 1983:29-30.) This is also the analysis assumed in Moser and Moser 1965. (I ignore here the internal structure of these syllables.)

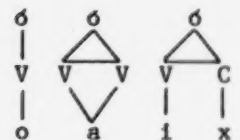
(11) a.



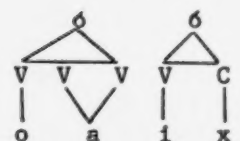
b.



c.



d.



I do not base this analysis primarily on the phonetic data, although it is certainly true that a word such as *kai:* 'mature' is not perceived as bisyllabic, despite the stress on the *a* and the length of the vowel *i*. It is also true that vowel clusters are not pronounced on separate beats in Seri songs. But the facts which I have already presented above also support this analysis. First, a polysyllabic analysis such as (11b) would create problems for stress assignment. If *kai:* is two syllables, then stress should be assigned to the long vowel—but it is not. Second, if (11b) is correct, then generalization (8) is not true. Third, if (11b) is correct, then rule (10) is wrong. The facts which these rules and generalizations describe would have to be accounted for in more complicated ways. The other structures in (11) do not have these problems associated with them.

Other evidence supports (11a). The only position in which a vowel is not preceded by a consonant in Seri is (1) in word-initial position

and (2) in words such as **koan̩a** 'tarantula' and **k-oa:ix** (NOM-hang:free). The first group can be accounted for by the following constraint on underlying forms in Seri:

- (12) Syllables without onsets occur only word-initially.

The second group is automatically accounted for under my analysis of them, by being eliminated. If all contiguous vowels are caught up into the same syllable in Seri, as in (11a), then **koan̩a** has only two syllables and they both have onsets; they do not violate (12).

Suppose we assume instead that a syllable nucleus can contain only one V, as in (11c). Constraint (12) can no longer be stated. We must then allow for syllables without onsets in (a) word-initial syllables, (b) the syllable following the stressed syllable, and (c) the second syllable following the stressed syllable *if* the first syllable following the stressed syllable had no onset, etc. Constraint (12), whatever its place in the phonology of Seri, is clearly preferable.

4.3 A restriction on vowel clusters

Sequences of a high front vocalic segment followed by a distinct nonconsonantal segment are generally associated with the syllabic skeleton CV. By all available tests morphemes beginning with **yo**, **ya**, and **ye** act as consonant-initial roots. (See Marlett and Stemberger 1983 for various facts which are relevant here, such as suppletive allomorphy dependent on syllable structure and the rule deleting vowels before vowels.) Such morphemes are common, but morphemes beginning with the sequences **ia**, **io**, and **ie** are nonexistent, with one exception -- the root **-ia:m** [-ea:m] 'wide' (see below). Morphemes beginning with the sequence **if** (a vowel sequence, not a long **i**), are attested in Seri.¹¹

The negative morpheme structure condition can be stated as (13).

- (13) No morpheme begins with a vowel sequence, the first member of which is **i** and the second member of which is not **i**. Exception: **-ia:m**.

But as it turns out, this constraint can apparently be subsumed under a more general constraint on vowel sequences. The vowel sequences **ia**, **io**, and **ie** do not occur except across morpheme boundaries. (They do occur in such derived sequences as **ionam** 'his hat' (< **i-aonam** 3POSS-hat). The constraint is given as (14):

- (14) No morpheme contains a vowel sequence, the first member of which is **i** and the second member of which is not **i**. Exception: **-ia:m**.

The only problem is the cluster which is phonetically [ea]. A regular phonological rule converts **i** to **e** before **a**. This is seen happening when the prefix **i-** occurs before roots of certain types (see note 12). This rule could account for the phonetic shape of **-ia:m**, which behaves by

phonological tests as beginning with a non-low vowel. (That is, with respect to suppletive allomorphy and phonological rules, the verb does not begin with a low vowel, but with a non-low vowel.) There are other examples of [ea]. One is [eaš] 'eelgrass', which might actually be bimorphemic, underlyingly /i-aaš/, with the third person possessor prefix. In such a case it would not violate (14). Or it might be underlyingly /eaš/, and thus not violate (14). The word [sea:to] is a loanword from Spanish *chivo* 'goat'. If it is underlyingly /sea:to/, it does not violate (14). At any rate, such examples are very rare.

Other attested vowel clusters are given below. Besides these there are a few instances of geminate vowel clusters. Since these sound phonetically just like 'long' vowels, they are identifiable only by morphophonemic alternations (see notes 9 and 11). In the examples of word-medial clusters below, I have tried to exclude word-final vowels since there is a tendency to lengthen vowels in this position, making the contrast between long and short hard to discern. Since word-initial clusters are very rare,¹² the examples in (15) are all word-medial, except where noted. I exclude words from this list which are derived via a morphologically conditioned ablaut process, even though the cluster is still internal to the root.

(15) Monomorphemic vowel clusters

a. Two short vowels

ai	tamāix	'gray-colored clay'
ao	?a-onam	'hat' (< /?a-aonom/)
oa	koaft	'whirlwind'
oe	koept	'quail'
oi	t-oit	(RL-dance)

b. Long plus short

a:a	i-t-a:a	(OM-RL-call)
a:i	sa:ix	'wasp (sp.)'
a:o	t-a:oɪ	(RL-pleated)

c. Short plus long

ai:	i-t-ai:X	(OM-RL-go:from)
ee:	i-t-anēe:x	(OM-RL-roll:up)
oa:	t-oa:n	(RL-turbid)
oe:	i-t-apōe:x	(OM-RL-harvest:eelgrass(pl.))
oi:	t-moi:x	(RL-circular)
ii:	i-t-aknfi:Xo	(OM-RL-pour:out)

d. Short plus long plus short

oa:i	t-oa:ix	(RL-hang:free)
------	---------	----------------

5 Codas

5.1 Simple codas

Any consonant may appear in a word-final simple coda, except **y**, although **f** is rare. **x** and **ʔ** are extremely common in unstressed word-final syllables, whereas **p**, **k**, and **s** are rather rare.

(16) Simple codas

Stressed syllable

ʔap	'mule deer'
Xat	'hail'
tok	'there'
i:f	'his/her nose'
tis	'harpoon point'
a:ʃ	'his/her grandfather'
a:ʔ	'his/her spouse'
ʃi:x	'plant (sp.)'
ʃi:X	'thing'
am	'her father'
ye:n	'his/her face'
tʃiʔ	'fish (sp.)'

Unstressed syllable

-aXop	'carry on pole'
simet	'bread'
iʔik	'seed, pit'
iyas	'his/her liver'
XomXʃiʃ	'plant (sp.)'
ʃi:ʃiʔ	'plant (sp.)'
ʃamix	'palm (sp.)'
ʔakiX	'where'
yaʃim	'its fin'
moXon	'fish (sp.)'
-aʔ	(focus)

The only example in which the glide **y** occurs in a coda is the loanword **ka:y** 'horse' (which is homophonous with the native word **k-a:i** (NOM/OM-make)). Except for this example, it is clear that whenever a vocalic segment occurs in a coda, it is attached to a V position in the syllabic skeleton rather than to a C position. The evidence for this comes from an insertion rule which applies, for example, when the declarative clitic **-ʔa** follows a consonant. Some examples without insertion are given in (17a), and examples with insertion are given in (17b). The examples in (17c) show that the morphemes in question do not have a final C.

(17)	a. V + ʔa	po:	po:-ʔa	'wild pig'
		ke:	ke:-ʔa	'bird (sp.)'

b. C + ?a	?ap	?ap-i?a	'mule deer'
	ʃi:k	ʃi:k-i?a	'bird'
	ka:W	ka:W-i?a	'plant (sp.)'
	ʃa:ʔ	ʃa:ʔ-i?a	'sun'
	ka:y	ka:y-i?a	'horse'
c. Vi + ?a	?ai	?ai-?a	'wind'
	?apā:i	?apā:i-?a	'Papago'
Vo + ?a	ka:o	ka:o-?a	(NOM-pass)

The restriction can be stated as in (18).

- (18) No coda has a C position associated with a vocalic segment. (Exception: **ka:y** 'horse')

Given this constraint, we have a reason for analyzing long vowels as a single segment attached to two V positions and not as a segment attached to a V and a C position. The latter would be in violation of constraint (18).

5.2 Complex codas

Complex codas are extremely common, although they are often derived by means of a rule of Syncope (discussed in Sect. 5.3.) Consonant sequences in the coda generally follow the universal sonority scale; thus obstruent-nasal and obstruent-glide clusters do not occur. There are two exceptions to this scale in Seri codas. As in onsets, clusters of obstruents in codas are freely permitted. In addition, some sonorant-sonorant clusters are permitted. Sequences of sonorant followed by obstruent are permitted by the sonority scale and also occur. Notice that **f** does not occur as a noninitial member of a complex coda. (Glosses to the words in (19) are included in the appendix.)

- (19) Word-final two-consonant codas

a. Stop as second consonant

Stressed syllable

	p	t	k
p	*	-Xopt	-Xto:pk
t	-atp	*	-i:tk
k	-ka:kp	pta:kt	*
f	-oafp	koaft	-a:fk
s		?ast	-no:sk
ʃ	Xto:ʃp	koʃt	?aXʃ:ʃk
ʔ	-iʔp		-a:iʔk
x	-koxp		itaxk
X	-o:Xp	a:Xt	-iXk

m	itāmt	akō:mk
n	?ant	?onk
?	-okē?t	

Unstressed syllable: ko:taik -a:taxk

b. Fricative as second segment

Stressed syllable

	s	š	ɬ	x	X
p	kops	-Xapš	?apɬ	-yanōpx	ksipX
t	-o:ts	aitš	-atɬ	-atx	šatX
k	taks	šakš	-atōkɬ	sa:kx	?akX
f		-kafš		-ifx	
s	*		-masɬ	Xa:sx	-asX
š		*	-apXāšɬ	šo:šx	-kašX
ɬ			*		?aɬX
x				*	
X		?aXš	-iXɬ		*
m	i?fms	iki?fmš	simɬ	-atā:mx	-amX
n	-tins	inš	-?anɬ	ya:nx	
?	i?s	-ifā?š		-amf?x	

Unstressed syllable:

?apats -me:sotš -a:katx Xpasf:tikɬ

As in onsets, no geminate consonants are allowed. Also unattested morpheme-internally is a sequence of the velar fricative and another fricative. A surprising fact is that **m** cannot occur in a complex coda (or onset) with **p**. This prohibition extends also to three-consonant codas (see below). The cluster **np** does not occur in two-consonant codas either, but it does in three-consonant codas.

Unlike in onsets, however, nasal-nasal clusters (specifically **nn**) occur in codas, although these are not common. Two examples are **ptkann** 'lobster' and **knōtōnn** (NOM-weak). Since these are violations of the sonority hierarchy, they must be explicitly allowed for.

Three consonant clusters in codas are also fairly common.

- (20) Word-final three-consonant codas (all are in stressed syllables unless otherwise noted)

a. Stop-Stop-Fricative

ptx	-aptx	'wide'
ptX	-a:ptX	'gather'
tkš	atkš	'her younger sister'
ktš	-aktš	'sift'

Unstressed syllable: ktx -akšesaktx 'prepare
wood for burning (pl.)'

b. Stop-Fricative-Stop

p+k	-aššp+k	'paralyzed in legs'
t+k	-Xat+k	'flat and thin (pl.)'
k+k	-apšk+k	'take shell off turtle (mult.)'
pxk	-akšpxk	'hurry to do'
txk	ka:txk	'grasshoppers'
pXt	-a:pXt	'gather eelgrass'

c. Stop-Fricative-Fricative

pXš	-kapXš	'sour'
tXš	-itXš	'break into pieces'
psš	ke:psš	'crab (sp.)'
tsx	-matsx	'brag'
ksx	-iksx	'immature (fruit)'
ksX	-aksX	'awake'
pšx	sšenšpšx	'blue heron'
pšX	-apšX	'torn out'

d. Fricative-Stop-Fricative

spX	-aspX	'winnow'
XpX	-aXpX	'mad at'
ftx	-noftx	'curved'
stx	istx	'its leaf'
škš	aškš	'his younger brother'
škš	-a?iškš	'gather fish testes'

e. Fricative-Fricative-Stop

f+k	-af+k	'knee'
š+k	-i?išk+k	'dirty'
X+k	-i:X+k	'convex (pl.)'
sxx	-isxx	'have body lice'
šxx	-ašxx	'leave piled up'

f. Fricative-Fricative-Fricative

fšX	-a:fšX	'fast'
šXš	-a:šXš	'cough'

g. Fricative-Stop-Stop

skt	-akδeskt	'prepare wood for burning'
škt	-oiškt	'go (pl.)'

h. Sonorant-Stop-Fricative

npX	-a:npX	'return home'
mtx	-ašāmtx	'make adobes'
ntš	entš	'knives'
ntX	?antX	'base'
nkɿ	-ankɿ	'pitted'

i. Sonorant-Fricative-Fricative

nɿX	-inɿX	'return empty-handed'
mšx	?amW i?ikδ:mšx	'tape worms'
nšX	-anšX	'spin thread'
?šX	-a:?šX	'sneeze'

j. Sonorant-Fricative-Stop

mɿk	-atāmɿk	'notched (pl.)'
nɿk	-aXδnɿk	'loosen (pl.)'
mxk	-amxk	'take, bring'
nxk	-monxk	'wavy'
?ɿk	-amɿ?ɿk	'plane smooth (pl.)'

Unstressed syllable: mxk ataɿXfkomxk 'his
brothers-in-law'

Some, but not all, of the gaps are accounted for by previously stated conditions. It is not clear whether any of the remaining gaps (such as the lack of three-consonant clusters ending in p) are particularly interesting, however.

5.3 Round Consonants in Codas

Number in nouns and verbs is often indicated by a change in the form of the root, either in addition to or in lieu of an affix. One of the most common root changes is the loss of the vowel in the post-tonic syllable. The environment for the rule today is clearly morphological, and very idiosyncratically so. Some examples showing this rule, which I call Syncope, are given in (21).¹³ Notice that Syncope sometimes applies in the derivation of the singular form, sometimes in that of the plural.

(21)	singular	plural	UR of root
	se:poɬ	se:pɬ-ox 'bush (sp.)'	/se:poɬ/
	moxépe	moxép-tox 'saguaro cactus'	/moxépe/
	?ast	?asat-ox 'stone'	/?asat/
	?apɬ	?apaɬ-ox 'tongue (abs.)'	/?apaɬ/
	o:t	o:to-ɪka 'coyote'	/o:to/
	Xti:p	Xti:po-ɪk 'shellfish (sp.)'	/Xti:po/
	koepɬ	koepit-ox 'quail (sp.)'	/koepit/

When an *o* is the target of the Syncope rule and there is a back consonant (*k*, *X*, or *x*) in its immediate environment, the *o* is not lost, although the *V* position is. Rather, the *o* remains and attaches to the back consonant, giving rise to the round consonants. The distribution of round consonants is therefore understandably extremely restricted. A few examples are given below. (Other morphological changes are ignored.)

(22) $\begin{array}{cccc} C & V & C & V & C \\ | & | & | & | & | \\ X & e & k & o & \dot{\tau} \end{array} \rightarrow \begin{array}{cccc} C & V & C & C \\ | & | & | & | \\ X & e & k & o & \dot{\tau} \end{array} \quad \text{Xek}^W\dot{\tau} \text{ 'wolves'}$

$\begin{array}{ccccc} C & C & V & C & V \\ | & | & | & | & | \\ k & \dot{s} & a & X & o \end{array} \rightarrow \begin{array}{cccc} C & C & V & C \\ | & | & | & | \\ k & \dot{s} & a & X & o \end{array} \quad \text{k}\dot{s}aX^W \text{ (NOM-talk)}$

$\begin{array}{ccccc} C & V & V & C & V & C & C \\ | & | & | & | & | & | & | \\ k & a & ? & o & x & k & \end{array} \rightarrow \begin{array}{ccccc} C & V & V & C & C & C \\ | & | & | & | & | & | \\ k & a & ? & o & x & k \end{array} \quad \text{ka:}^?Wk \text{ (NOM-squat)}$

Round consonants in word-final codas include those shown in (23) (see the appendix for glosses). While most of them are derived (historically, at least) via Syncope, I do not claim that all of them are, although that may be the case.

(23) a. Final position, preceded by vowel:

ka:W
-ipaW
kpataW
-ak^W
ka:X^W

b. Final position, preceded by consonant:

-am^W
 -e:s^W
 inδ:ʔk^W
 -i:pk^W
 ktamk^W
 anX^W
 -isX^W
 -omtX^W

c. Penultimate position, preceded by vowel:

kakāWk [kakāWk^W]
 Xek^Wɿ
 -ik^Ws
 ʔaX^Wɿ

d. Penultimate position, preceded by consonant:

ka:ʔWk [ka:ʔWk^W]
 -i:pWk [-i:pWk^W]
 to:tWk [to:tWk^W]

e. Antepenultimate position:

-ik^Wsx

The rounding of the final **k** in words such as **kakāWk** in (23c) is correctly described by a late rule. In fact, I take this rounding as postlexical (in the theory of Lexical Phonology (Mohanan 1986)). A completely productive rule which (under conditions which have to do with the presence of certain adverbs) inserts **a** between two post-tonic consonants reveals that it is **k** rather than **k^W** when **a**-Epenthesis applies.

(24)	Word	After a -Epenthesis	
	ʔast	ʔasat	'stone'
	Xe:k ^W ɿ	Xe:k ^W aɿ	'wolves'
	kakāWk	kakāWak *kakāWak ^W	'bird (sp.)'
	ka:W	---	'bush (sp.)'
	kāaX ^W	---	(NOM-talk)

6 Word-medial consonant clusters

The word-medial consonant clusters in (25-28) are attested. Although I cannot claim that these lists are exhaustive, they are fairly complete and serve to illustrate the possibilities. Morpheme boundaries are not indicated here. The only morpheme boundaries which occur between the clusters in question are those which pertain to singular/plural and singular/multiple action morphology (see the

beginning of Sect. 5.3). This morphology is as complex for verbs as singular/plural morphology is for nouns; it is discussed in detail in Moser 1961 and Marlett 1981. The verbs in the list are cited in the subject nominalized form, although this is not indicated in the gloss. For example, the verb form **kiksim** in (26a) is underlyingly /k-i-akas-tim/. (The sequences which result from the introduction of prefixes are discussed in Sect. 8.1.) In the following lists I also indicate by means of "I" (for initial) or "F" (for final) those clusters which occur word-initially and word-finally, respectively, as established above (ignoring stress).

(25) Two consonant clusters, stop as second segment

Stop-Stop

Following a stressed vowel:

tp	F	ko:tpam	'sardine'
kp	I,F	katikpan	'work'
pt	I,F	kaptax	'open (mult.)'
kt	I,F	ki:ktō	'pregnant (mult.)'
pk	F	kapka	'rain'

Fricative-Stop

Following a stressed vowel:

fk	F	i:fkox	'noses'
sp	I	ko:spox	'spotted'
st	I,F	ka:stax	'spread'
sk	I,F	kaskim	'row'
ʃt	F	kaʃtaʃ	'wheat'
ʃk	I,F	kaʃkim	'go'
Xk	I,F	kiXkim	'throw'

Following an unstressed vowel:

Xt	I,F	kiXtō:ʃ	'scar'
ɬk	F	ka:kaɬka	'dress (mult.)'
xk	I,F	ka:taxkim	'provoke'

Sonorant-Stop

Following a stressed vowel:

mk	F	kimkaxk	'want (pl.)'
?t	F	ki:ʔtox	'be (pl.)'
?k		kaʔka	'exist'

(26) Two consonant clusters, fricative as second segment

Stop-Fricative

Following a stressed vowel:

ks	I, F	kiksim	'chew'
pš	F	katšpša	'urinate (woman) (pl.)'
tš	I, F	ka:otšit	'choke (pl.)'
pž	F	kopžim	'bird (sp.)'
tž	F	katža	'be afraid'
px	F	kakšpxom	'shellfish (sp.)'
pX	I, F	kipXox	'blow'
tX	F	katXo	'many'

Following an unstressed vowel:

kž	F	kažfšakžox	'gather fish testes (pl.)'
----	---	------------	----------------------------

Fricative-Fricative

Following a stressed vowel:

Xš	F	kkoXša	'babysit (pl.)'
sx	F	kasxox	'row (pl.)'
šx	F	kkašxa	'bite (pl.)'
šž	F	kišžax	'leave piled up (mult.)'
šX	F	ka:išXa	'bring, take (pl.)'

Following an unstressed vowel:

sž	I, F	kasžšepk	'delouse (head lice)'
sX	F	isXšp	'top of his/her head'
žX	I, F	atažXf:kom	'his brother-in-law'

Sonorant-Fricative

Following a stressed vowel:

mš	F	ki:mšo	'want'
nX		kakš:nXot	'help carry'

(27) Two consonant clusters, sonorant as second segment

Stop-Sonorant

Following a stressed vowel:

pn	I	kopni	'paper wasp'
tn		kitni	'touch against'
kn	I	kakni	'bird (sp.)'
py	I	kakšpyas	'talk jerkily'

ty	kʔatyaxk	'stretch out (pl.)'
ky I	yekyar	'cowboy'

Following an unstressed vowel:

tm I	atmǎʔax	'his/her aunt'
km I	akmǎʔax	'his niece/nephew'

Fricative-Sonorant

Following a stressed vowel:

fn	i:fnix	'his/her nostril'
ǃn I	kka:ǃni	'bite'

Following an unstressed vowel:

sm	kasmǎnexa	'make arrow shaft'
ǃm	aǃmǎ:	'his/her uncle'
sn I	kaenǎ:pít	'disturb'

Sonorant-Sonorant

Following a stressed vowel:

ʔn	ka:ʔnix	'tremble'
----	---------	-----------

(28) Word-medial three-consonant clusters

Following a stressed vowel:

ǃk F	kaʔiǃkox	'tie up'
tXǃ F	kkatXǃa	'bite'
sǃk F	kko:sǃka	'sprinkle'
tǃk F	kamǎXatǃka	'cause to be a drab color (pl., mult.)'
kǃk F	kapǃ:kǃkam	'look behind (pl.)'
tXt	kkotXta	'malodorous'
pXǃ F	ka:pXǃoj	'gather eelgrass (pl.)'
XpX F	kiXpXax	'quiver'
ǃXǃ F	kiǃXǃim	'cause rash'
k ^w sX	kak ^w sXax	'long, tall'
k ^w ǃx	ʔayǎk ^w ǃxox	'anklebones (abs.)'

Following an unstressed vowel:

tkm	atkmǎǎ:m	'her brother-in-law'
Xǃk F	kkaskaXǃka	'share (pl.)'
ʔxk	kaʔxkǎiǃ	'fall (pl.)'

Given the fact that Seri has both syllable-initial and syllable-final clusters, established by their occurrence word-intially

and word-finally, it may be asked how word-medial clusters are syllabified. It is obvious that these sequences cannot be wholly identified either with those which occur syllable-initially or those which occur syllable-finally. I assume that the syllable division is between the two consonants.

The syllabification of three-consonant clusters is more difficult, however. In *kaʔfɛkox*, the first example in (28), is the syllable division between the *ʔ* and the *k*, or is it between the *k* and the *x*? The distribution of the consonants to this point does not seem to provide any help. There is, however, one important fact which may be used to argue for one syllabification over another, and it also indicates that syllabification is stress-sensitive in Seri: The phoneme *ɲ* has a lenis allophone of nasalized *w* in one specific situation. Alternation evidence supports this analysis. Compare the allomorphs of the negative prefix in the following words:

- (29)
- | | | | |
|----|--------------|---------------------|-----------------|
| a. | t-ataX | 'did he go?' | (RL-go) |
| | t-[m]-ataX | 'didn't he go?' | (RL-NEG-go) |
| b. | yo:taX | 'he went' | (DIST-go) |
| | yo-[m]-ataX | 'he didn't go' | (DIST-NEG-go) |
| c. | i-t-aʔo | 'did he see it?' | (OM-RL-see) |
| | i-tk-[w]-aʔo | 'didn't he see it?' | (OM-RL-NEG-see) |

The lenis allophone also occurs in the words in (30).

- (30)
- | | | |
|---------|----------|----------------|
| kma:m | k[w]a:m | 'woman' |
| kmaxf:k | k[w]xf:k | 'women' |
| kmi ke | k[w]i ke | 'Seri, person' |

This allophone does not occur in the word in (31), however. The allophone [m] appears.

- (31) i-t-ákmox itak[m]ox (OM-RL-put:long:thing(pl.))

Assuming that the *k* and *ɲ* belong to two separate syllables in the word *itákmox*, I suggest that the phonetic rule leniting *ɲ* must refer to a tautosyllabic velar stop. But if this is true, then the syllabification of the word in (32), as shown, is different from that in (31) because the lenis allophone occurs in the word in (32).

- (32)
- | | |
|--------------------|-----------|
| ak[w]áʔax | [+stress] |
| 'his nephew/niece' | |
| | |

In the case of a three consonant cluster in (33), the *k* must be syllabified with the *ɲ* since the lenis allophone occurs.

- (33) atk[^W]axé:m [+stress]
 'her brother-in-law'
- $\begin{array}{ccc} \triangle & \triangle & \triangle \\ a & t & kma \quad xe:m \end{array}$

The conclusion I draw from these facts is that the medial sequence **tkm** is syllabified somewhere before the **k**, and that the medial sequence **VkmV** is syllabified immediately before the **m** if the second vowel is unstressed and immediately before the **k** if it is stressed. Speaking more generally, then, onsets in Seri are maximized (to some limit perhaps) in stressed syllables but not in unstressed syllables.

7 Syllable overview: Uninflected words

Drawing on the facts discussed above, we can make certain claims about syllables in words which do not involve affixation. It appears that it is important to analyze word-initial and word-final syllables separately from word-medial syllables, and stressed syllables from unstressed syllables.

The Onset of a Seri syllable usually has one or two consonants. In word-medial position, it has at least one. The two word-initial cases in which the onset has three are not covered by (34).

- (34) Onset: C_0^2 If word-medial: at least one

The Nucleus of a Seri syllable has between 1 and 3 vowel positions. Multiple vowel positions are permitted only in stressed syllables.

- (35) Nucleus: V_1^3 If unstressed: only one

The coda of a Seri syllable has between 0 and 3 consonants. Word-medially the maximum is one.

- (36) Coda: C_0^3 If word-medial: at most one

8 Derived words

Affixation in Seri results in a number of consonant clusters which do not occur morpheme-internally. A significant number of these clusters are allowed to surface, but sometimes phonological rules apply to prevent them from occurring. In this section I discuss both the new consonant clusters and also the rules which 'fix up' a cluster which is not permitted.

8.1 New clusters

There are several affixes which, upon attaching to verbs, interact with the previously discussed constraints on Seri syllable structure in interesting ways because they consist of consonants only or have allomorphs consisting of a single consonant. These affixes (whose

allomorphs shown do not include changes in point of articulation) are given in (37). (Noun morphology is not included here since those affixes do not interact in any interesting way.)

(37)	Prefixes	si- ~ s-	IRR	m-	2sSUB
		mi- ~ m-	PROX	m-	NEG
		ko- ~ k ^w -	3OBL	tm-	ABIL
		t-	RL	?-	1sSUB (transitive)
		k-	NOM	?p-	1sSUB (intransitive)

Infix -ko- PL

Suffix -X UT

The first observation is that affixes not involving a sonorant consonant attach freely to almost any syllable (syllables with sonorants are discussed in Sect. 8.2-4 below). This gives syllable onsets and codas which are one or two segments longer than those which occur in the root. Identical consonants may be juxtaposed by this means.¹⁴ Some of these are given in (38) in the surface form. (The pronunciation of a cluster such as *tt*, whether word-initial or word-medial, is as a long *t* -- that is, as a *t* with the closure prolonged.)

(38) Four-consonant onset: k^w-t-ʃXok (3OBL-RL-hack:off)

Three-consonant onset: k-ʃXok (NOM-hack:off)
 t-Xtamt (RL-abundant)
 k-pko:yo (NOM-taste/PL)
 (< root -pi: + infix -ko-)

Complex onset with ? : k-?e:t (NOM-red)

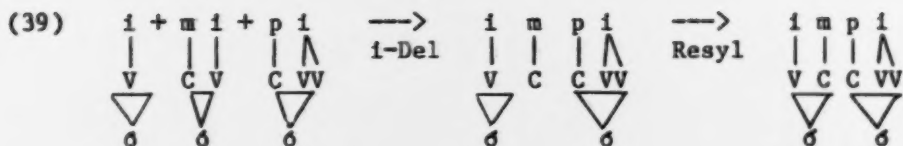
Complex onset with f : k-fikx (NOM-cover)

Identical consonants juxtaposed: t-toknix (RL-round)
 k-kap (NOM-fly)
 s-samix (IRR-rolled:up)

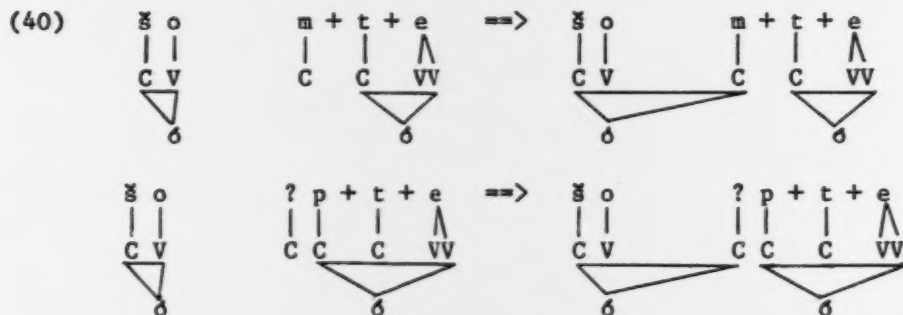
Four-consonant coda: i-t-amxk-X (OM-RL-bring-UT)

8.2 Resyllabification

The *i* of the prefixes *i-*, *mi-* and *si-* is lost under certain conditions, including before consonants (see Marlett and Stemberger 1983). When this happens, the *s* attaches to the onset of the following syllable, but the *m* cannot, since such an adjunction would be in violation of the sonority hierarchy used by Seri. In some cases, this creates no problem since it can form the coda of the preceding syllable, as the derivation in (39) of *impi*: 's/he tasted it' (OM-PROX-taste) shows.

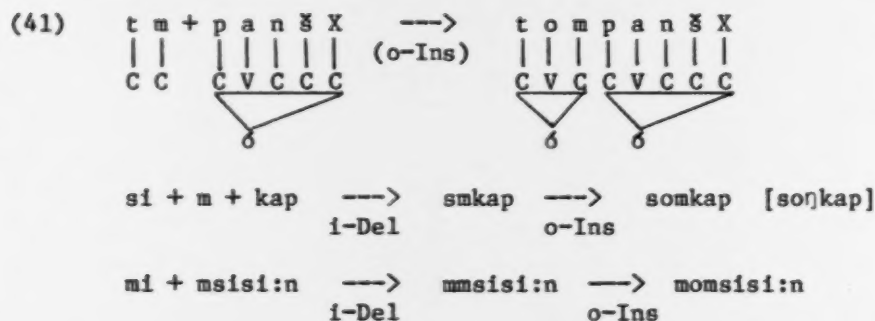


Likewise, an unattached sonorant which results directly from affixation may attach itself to the preceding word if the sonorant occurs word-initially. This is illustrated in (41) for **šo mte**: 'what (lit. 'how') did you say?' (2sSUB-RL-say) and **šo ?pte**: 'what did I say?' (1sSUB-RL-say).



8.3 o-insertion

In some cases an **o** is inserted before an unattached word-medial sonorant (which happens to always be a nasal). This is illustrated in (41) for **toṃpanšX** 's/he can run' (ABIL-run), **soṃkap** 's/he will not fly' (IRR-NEG-fly'), and **moṃsisi:n** 's/he is pitiable/loveable' (PROX-pitiable).



The rule is given as (42). The condition included on the rule prevents the rule from overapplying; an **o** is not inserted in this position if the first consonant is part of a person prefix. This condition could be dispensed with by specifying the levels at which the rule applies, as is done in Lexical Phonology. (I use the notation of a bar above the consonant to indicate that it is not attached to any syllable.)

Appendix: Glosses of examples not glossed in text

-a:fk	'hit hard'
-a:iɬk	'undulate (mult.)'
-a:katx	'put out'
-ak ^w	'kill'
-a:ɬaxk	'leave behind (pl.)'
-amf?x	'plane smooth'
-amX	'say'
-amW	'anus'
-apXáɬɬ	'run into'
-asX	'count'
-atá:mx	'notched'
-atɬ	'afraid (pl.)'
-atɔkɬ	'cut (pl.)'
-atp	'spit'
-atx	'trunk of body'
-e:sW	'barefoot (archaic)'
-ifá?ɬ	'sob'
-ifx	'have indigestion'
-ik ^w s	'spiny'
-ik ^w sx	'shake out'
-iɬp	'sniff'
-ipaW	'straighten (mult.)'
-i:pk ^w	'make pounding sound'
-i:pWk	'make pounding sound (pl.)'
-isX ^w	'hide'
-i:tk	'drip'
-iXk	'melt (pl.)'
-iXɬ	'grab'
-kafɬ	'snap at'
-ka:kp	'hit to remove what's inside'
-kaɬX	'break (intrans.)'
-koxp	'bolt, jump'
-masɬ	'yellow (pl.)'
-me:sotɬ	'untouched (pl.)'
-no:sk	'toothed'
-oafp	'bounce off'
-oké?t	'bouncy'
-omtX ^w	'straight'
-o:taɬk	'ants'
-o:ts	'hiss'
-o:Xp	'white'
-tins	'scrape'

-Xapš	'congeal'
-Xopt	'extinguished'
-Xto:pk	'stay overnight'
-yanōpx	'fist'
-?an±	'ten'
aitš	'his/her uncle'
akō:mk	'his younger sister'
a:mo	'far'
anX ^w	'much'
asa:k	'his son'
ašō:k	'her nephew/niece'
ašox	'star'
a:Xt	'plant (sp.)'
fe:±	'duck (sp.)'
i:fa	'peninsula'
i:kx	'coarse sand'
iki?imš	'ringworm'
i±ft	'his/her head/hair'
inō:?k ^w	'bay'
inš	'his/her spinal cord'
i:stoX	'their breath, soul'
itāmt	'his/her sandal'
itaxk	'his/her bones'
iti	'on'
iXox	'cactus fruit (stage)'
i?ims	'fringe'
i?s	(part of idiom)
kakāWk	'bird (sp.)'
kasāmikt	(NOM-jealous)
ka:X ^w	'belonging to another person'
ka:W	'bush (sp.)'
ka?a	'fish (sp.)'
ka:?Wk	(NOM-squat)
ke:±X	'paddle'
kma:m	'woman'
kmaxi:k	'women'
knaxox	(place name)
koaft	'whirlwind'
kofikx	(NOM-D-cover)
konē:	'grass (sp.)'
kops	'glowworm'
košt	'cricket'
kotōpis	'shellfish (sp.)'
ko:ta±k	'ants'
ko:xox	(NOM-short)
koyōko	'dove'

kpataW	'fish (sp.)'
kpo:t	'fish (sp.)'
ksai	'hairbrush'
ksipX	'glue'
ktam	'man'
ktamk ^w	'men'
ktamδ:	(NOM-fierce)
kyaXa	'saguaro cactus (stage)'
ʔanā:kx	(place name)
ʔXe:kox	(place name)
mo:snl	'sea turtle'
moXfma	'yesterday'
naxδ:	'fish (sp.)'
no:nl	'butterfly larva (sp.)'
nop	'mountain lion'
pak	'some'
paʔit	'head of deer for dancing or hunting'
paxā:s	'squid'
pna:kox	'mangrove'
potā:	'larva'
psa:k	'hunger'
pta:kt	'plant (sp.)'
pXa:ʔom	'three (counting)'
pyoke	'starfish (sp.)'
sa:kx	'shellfish (sp.)'
sanā:kx	(place name)
set	'bird (sp.)'
simi	'barrel cactus (sp.)'
ska-	(first person plural imperative)
sʔe:	'heron (sp.)'
sʔe:kox	'heron (sp.)'
sna:šX	'plant (sp.)'
spitx	'plant (sp.)'
stak	'pumice'
šakā:m	'unmarried girl'
šakš	'shellfish (sp.)'
ša:p	'roadrunner'
šatX	'tiny thorns'
škaXt	'young (archaic)'
šnapX ^w ʔ	'plant (sp.)'
šo:šx	'burlap bag'
taks	'plant (sp.)'
tamāix	'gray-colored clay'
taxisʔ	'fish (sp.)'
tis	'harpoon point'
tm-	(abilitative)

to:tWk	'cholla cactus (sp.)'
tš1?	'fish (sp.)'
tšikáktox	'cardinal'
txamoxa	'fish (sp.)'
xkoa	'plant (sp.)'
xoene	'plant (sp.)'
Xapó:	'sea lion'
Xa:sx	'whale (sp.)'
Xašš:p	'peacock'
Xek ^W ž	'wolves'
Xe:že	'fog'
Xepe	'sea'
Xkokni	'cactus (sp.)'
Xžožo	'fish (sp.)'
Xna?š:iž	(place name)
Xnois	'dirt, refuse'
Xpakš:fk	'crab (sp.)'
Xpasf:tikž	'ground squirrel (sp.)'
Xpašx	'graphite'
Xpeyo	'sailfish'
Xtamš:ixa	'tortoise (sp.)'
Xti:p	'shellfish (sp.)'
Xto:šp	'plant (sp.)'
Xxi:	'gourd'
ya:nx	'poisonous sap'
yatš:m	'mountain pass'
ye:n	'his/her face'
?akX	'somewhere'
?akšmet	'small lizard (gen.)'
?ažX	'rather'
?amš:n	'family'
?ant	'land'
?apats	'Apache'
?apž	'tongue (abs.)'
?aso	'net'
?ast	'stone'
?aXš:šk	'arrows'
?aXš	'dog, pet'
?aX ^W ž	'shellfish (sp.)'
?a?š:t	'door'
?eme	'camp'
?epem	'whitetail deer'
?onk	'duck (sp.)'

Notes

*Seri is spoken in northwestern Mexico by about 600 people.

I thank Diana Archangeli, G. Hubert Matthews, Mary B. Moser, Joseph Stemberger and two anonymous IJAL reviewers for their valuable comments on an earlier draft of this paper. They are absolved of responsibility for remaining inadequacies. I especially profited from the use of the prepublication copy of the Seri dictionary which Mary B. Moser is currently preparing. This dictionary is based on data collected by Mrs. Moser and her late husband Edward Moser since 1951 under the auspices of the Summer Institute of Linguistics.

The following abbreviations are used: ABIL--abilitative, abs--absolute, Co--coda, D--detransitivizer for unspecified object, DIST--distal, gen--general term, IRR--irrealis, mult--multiple action, NEG--negative, NOM--nominalizer, Nu--nucleus, OM--object marker, pl, PL--plural, PROX--proximal, RL--realis, sp--species, UR--underlying representation, UT--unspecified time, 2sOBJ--second person singular direct object, 3OBL--third person oblique, 1sSUB--first person singular subject, 2sSUB--second person singular subject, σ--syllable.

1. Words with this voiced lateral include the following:

alo kik^Ws (dialectal variant of xto^Wo kik^Ws) 'mullet (sp.)'
 -lao 'gobble up'
 lams 'fish (sp.)' (archaic)
 laXf:ktim (place name)
 Xalá: 'bird (sp.)'
 Xpalémeik 'shellfish (sp.)'
 -oalaxk 'leave behind, detransitivized, pl.'
 -i:ola 'puffed out'
 -oalala 'flap'
 itXalóms 'short antenna of lobster'

With the exception of a few Spanish loanwords, this list may be exhaustive for current Seri.

2. Words with r include the following:

Seri	Spanish	English
ro:k ^W	loco	crazy
?o?ra	burro	donkey
ret	reata	lariat
kaskaréra	escalera	ladder
koksar		Mexican (origin unknown)
tear	diablo	devil
pa:r	padre	priest
-apré:nt	empeñar	lend
-apásirox	pasear	take a stroll, ride
matar		(plant name) (origin unknown)

merɔ:n	melɔn	melon
siiɾ	silla	saddle
o:ro	oro	gold, money, dollars
to:tar		chicken (origin Uto-Aztec)
yekyar	vaquero	cowboy

3. While the discussion in Sect. 5.3 suggests that **W** may in fact be the rounded velar fricative **x^W** phonologically (although not phonetically), I will continue the transcription **W** in the tradition of previous publications on Seri phonology.

4. Moser and Moser (1965) included nasalization and **as** phonemes. These are both derived from underlying **ɱ** in my analysis.

5. The underlying form of **itak** is /i-atak/, and that of **i:stoX** is /i-astoX/.

6. The word **e:nim** 'knife' is the only example of a monomorphemic bisyllabic word beginning with stressed **e**. It is a loanword, probably cognate with Papago **wainomi** 'metal'.

7. Moser and Moser (1965:62) report that in some cases there is open transition between consonants:

Voiced open transition occurs between members of the sequences **?t mk mn nk nx** when these sequences occur immediately following a stressed vowel: **ki?t** [kiʔ^ht^h] 'to hiccough'. With some speakers, close transition occurs in all other types of consonant sequences. With still other speakers, voiceless open transition occurs in sequences consisting of stop plus stop (except **? plus other stop**) and stop plus nasal. Close transition occurs in geminate clusters.

It must be emphasized, however, that any such open transition is extremely slight. The articulation of clusters is generally very close. I know of no differences between the articulation of common clusters and less common clusters. Finally, I wish to make clear that consonants in these clusters (and others in the language, except the one discussed in Sect. 8.4) are not syllabic.

8. If we were to specify the universal sonority scale in more detail, mentioning manner or place of articulation, Seri contains a large number of marked onsets and codas.

9. I do not mean to say that there are no sequences of identical vowels, since I do want to allow for such sequences when the length is extra long. The verb root **-sii:x** which occurs in the verb **-sii:x-im** 'move' has two vowels phonologically, although it is simply one long vowel phonetically. The verb undergoes an ablaut rule to form the plural stem, which is **-sia:i-am**. Similar facts (although in the reverse) are observed in the verb 'wash one's hands', which is **-onfi:-X** in the singular and **-onfi-ta-x-k** in the plural. The infixation of **-tɔ-**

in the formation of some plural forms also gives evidence for geminate vowel clusters. Compare **-i:ŋ** 'go (sg.)' and **-i-tō-i:x** 'go (pl.)'; **-a:a** 'call (sg.)' and **-a:-tō-a-t** 'call (pl.)'. Similarly, the verbs **-ai:X** 'leave' and **-ai** 'tell', which begin with a short low vowel and enter into the morphophonemic alternations which are typical of such roots (see (6)), undergo an ablaut rule to form the plural stems, which are **-aa:X-ox** and **-aa:-m**, respectively. The latter forms, although they have very long vowels phonetically, also enter into the morphophonemic alternations which are typical of roots beginning with short low vowels. The verb **-aa:ɪ** 'order' is another example. I know of only two verbs which demonstrably contain a sequence of short identical vowels phonologically: **-aa** 'know' and **-ee** 'give'. Compare the paradigms below with those in (6-7):

k-iy-a	(NOM-OM-know)	k-iy-e	(NOM-OM-give)
i-t-aa	(OM-RL-know)	i-t-ee	(OM-RL-give)
i-po-a	(OM-IRR-know)	i-po-e	(OM-IRR-give)
i-siy-a	(OM-IRR-know)	i-siy-e	(OM-IRR-give)

10. Stress is sometimes (but not always) perceived as being on the second vowel if the first vowel is nonlow. For example, in **koanta** the **o** has less stress than **a**. I take this as resulting from a phonetic gliding rule and not relevant to the matter of primary stress assignment.

11. Roots with the sequence **ii** include **-akni:xo** 'pour out', whose plural form is **-akn[e]a:ɪim**, and **-a?xfit** 'fall', whose plural form is **-a?x[e]a:tim**. Pluralization of these, and other, stems involves the ablaut of the second **i** to **a**.

12. Monomorphemic word-initial vowel clusters are almost nonexistent. The word [eaʃ] may be one monomorphemic example, although the actual vowels involved suggest a possible bimorphemic source historically, as noted above. This is the case for other attested examples of word-initial vowel clusters. For example, the third person possessive prefix **i-** on the root **-aa:x** 'vertebra' results in a word-initial vowel cluster: [ea:x]. The imperative form of intransitive verbs whose roots begin with **a** or **o** take a zero allomorph of the imperative morpheme, plus ablaut (if applicable). The imperative form of **-otx** 'arise' is **atx**, for example. An intransitive verb root beginning with a vowel cluster will therefore display an initial vowel cluster in the imperative form: **-oit** 'dance', **ait** 'Dance!'.

13. This rule is obviously what historically produced the large number of complex codas. Moser and Moser (1965) noted that, given the number of segments which can potentially occur in the onset, nucleus, and coda of a syllable, one might expect to find syllables of the form CCCCVVVVCCCC, which nevertheless do not occur. They attribute the nonoccurrence of such syllables to a matter of syllable weight. They state (p. 52), "A heavy build-up of the pre-vocalic margin precludes a like build-up of the post-vocalic margin and vice-versa. Likewise, an expanding vowel nucleus is paralleled by a proportionately decreasing content of the

then could be stated simply as shown below, applying only to enclitics. The rule does not need to specify the content of the V which is inserted.

Enclitic V Insertion:

$$\begin{array}{ccc} C & C & \\ | & | & \\ \alpha & ? & \end{array} \Rightarrow \begin{array}{ccc} C & V & C \\ | & & | \\ \alpha & & ? \end{array}$$

References

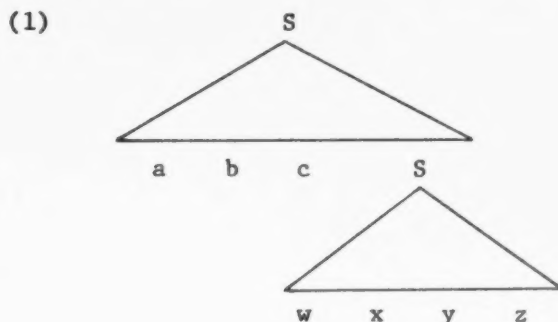
- Clements, George N. and Samuel Jay Keyser. 1983. CV phonology: A generative theory of the syllable. Linguistic Inquiry monograph 9. Cambridge, MA: MIT Press.
- Cole, Jennifer. 1987. The interaction of phonology and morphology in Seri, paper presented to NELS.
- Harris, James W. 1983. Syllable structure and stress in Spanish: A nonlinear analysis. Linguistic Inquiry monograph 8. Cambridge, MA: MIT Press.
- Marlett, Stephen A. 1981. The structure of Seri. Unpublished doctoral dissertation, University of California, San Diego.
- Marlett, Stephen A. 1984. Personal and impersonal passives in Seri, in Studies in relational grammar 2, ed. by David M. Perlmutter and Carol G. Rosen. Chicago: University of Chicago Press.
- Marlett, Stephen A. and Joseph P. Stemberger. 1983. Empty consonants in Seri, Linguistic Inquiry 14:617-639.
- Mohanan, K.P. 1986. The theory of lexical phonology. Dordrecht: D. Reidel.
- Moser, Edward W. 1961. Number in Seri verbs. Unpublished M.A. thesis, University of Pennsylvania.
- Moser, Edward W. and Mary B. Moser. 1965. Consonant vowel balance in Seri (Hokan) syllables, Linguistics 16:50-67.
- Selkirk, Elizabeth. 1984. On the major class features and syllable theory, Language sound structure, edited by Mark Aronoff and Richard T. Oehrle. Cambridge, MA: MIT Press.

EXTRACTION FROM COMPLEMENT CLAUSES IN KOINE GREEK*

Stephen A. Marlett

It is evident that, despite a certain amount of word order freedom in Ancient Greek, there are also severe constraints on where words may occur. In this brief paper I present one such constraint and show how it affects the understanding of two verses which seem to be counterexamples to it (1 Cor 14:12 and 1 John 5:16).¹

Dependents of a clause in Greek may not be freely interspersed with dependents of another clause, just as they cannot be in English. Therefore, a sentence with the structure shown in (1) (in which a sentence (S) is embedded inside of another sentence) cannot have the surface form shown in (2) (among others).



(2) a b w x c y z

There are, however, three ways in which this intuitively correct constraint can be superficially violated. The first is when an interrogative word of an embedded clause occurs sentence-initially, as in (3-4). This is referred to as WH-Movement in the transformational linguistic literature.² (Clause boundaries are indicated by square brackets; the (probable) source position of a displaced constituent is indicated and the displaced constituent is in boldface.)

(3) **What** do you think [he will do ___ for you] ?

(4) **ti** thelete [poiēso ___ hymin]
 what you.want I.may.do to.you
 'What do you want me to do for you?' Matt 20:32 (NIV)

The second is when a constituent of an embedded clause is allowed to become a constituent syntactically of the matrix clause; this is not a matter of simple movement. In English this 'raising' (as it is known in the earlier transformational linguistic literature) requires an infinitive

tive in the embedded clause, as it could in Greek. See examples (5-6). But in Koine Greek (as in Modern Greek, certain North American Indian languages, and apparently even Hebrew) raising can also occur with finite complement clauses and a pronominal 'copy', as in examples (7-8).³ In both kinds of raising, the case of the raised nominal is determined by its function in the higher clause. (Raising is not obligatory, and it is not clear what functional significance it might have had in Greek. Very possibly the relationship between form and function was just as slippery as in English.) According to this analysis, the pronoun **se** in (6) occurs in the accusative case since it is the direct object of the preceding verb at the surface level. Likewise, the noun phrase **to euaggelion** 'the gospel' in (7) is the superficial direct object of **gnōrizō** 'I make known' by this analysis, and the first pronoun **heautous** 'yourselves' in (8), a reflexive copy of the second one, is in the accusative case because it is the direct object of the preceding verb.

- (5) a. I expect **John** [____ to leave soon].
 b. **John** is expected [____ to leave soon].
- (6) **euchomai se** [____ **euodousthai kai hygiainein**]
 I.pray you to.prosper and to.be.in.health
 'I pray that you may prosper and be in good health'
 3 John 2
- (7) **gnōrizō gar hymin adelphoi to euaggelion ...**
 I.make.known for to.you brothers the gospel
 [**hoti** ____ **ouk estin kata anthropōn**]
 that not it.is according.to man
 'I would have you know, brothers, that the gospel ...
 is not of human origin' Gal 1:11
- (8) **he ouk epiginōskete heautous** [**hoti Iēsous**
 or not you.know yourselves that Jesus
Christos en hymin]
 Christ in you
 'Or do you not understand that Christ Jesus is in
 you...?' 2 Cor 13:5

Other clear examples of raising include Matt 25:24, Mark 1:24, Mark 7:2, Mark 11:32, Mark 12:34, to name a few. In my opinion, the construction in Matt 6:28 and Luke 12:24 should also be taken as raising, but a serious discussion of raising would take us far afield now.

Since raising involves a grammatical change which is not simply a change of word order, it is quite different from both WH-Movement and the construction which is the topic of this article: extraction from various kinds of finite subordinate clauses.⁴

Extraction in Greek is different from each of the constructions mentioned above. It simply involves the placement of a constituent in the position immediately preceding the complementizer, presumably for some kind of highlighting effect. In this position it looks like it could be part of the preceding clause, but I assume here that the moved element is adjoined to the complementizer. This is shown diagrammatically in (9).

(9) [w x [COMP y a z]] ==> [w x [[a COMP] y z]]

In other words, except for cases of WH-Movement, the only place in which an element that is uniquely the constituent of a finite subordinate clause may appear outside of that clause is immediately preceding the complementizer. In the absence of native speakers of Koine Greek, the validity of this constraint can only be extrapolated from the texts we have and corroborated by the examination of additional textual material. It may be that there are less than eighteen examples of extraction from finite clauses in the NT. I present the better examples below.⁵

Subordinate Clause Extracted

(10) **kathōs ēgapēsa hymas** hina kai hymeis agapate
as I.loved you that also you you.may.love

allelōus —
one.another

'that you love each other as I have loved you' John 13:34

Prepositional Phrase Extracted

(11) **legōn eis ton erchomenon met auton** hina pisteusōsin —
saying in the coming(one) after him that they.may.believe

'telling [them] that they should believe in the one coming after him' Acts 19:4

Noun Phrase Extracted

(12) **egō de hoti — tēn alētheian legō**
I but because the truth I.say

'but because I tell the truth' John 8:45

- (13) **tois ptōchois** hina ti dō
to.the poor that something he.may.give —

'that he should give something to the poor' John 13:29
- (14) houtōs kai houtoi nyn ēpeithēsan tō hymeterō
so also these now they disobeyed by.the your

eleei hina kai autoi nyn eleethosin
mercy so.that also they now they.may.receive.mercy —

'so they too have now become disobedient in order that they too
may now receive mercy as a result of God's mercy to you'
Rom 11:31 (NIV)
- (15) **biōtika** men oun **kritēria** ean echēte
of.this.life indeed therefore lawsuits if you.have —

'therefore, if you have lawsuits about everyday matters'
1 Cor 6:4
- (16) **tini** logō euēγγελισαμέν hymin ei katechete
to.what word I.preached to.you if you.hold.fast —

'if you hold firmly to the word I preached to you' 1 Cor 15:2
(NIV)
- (17) alla **tēn agapēn** hina gnōte hēn echō ...
but the love so.that you.may.know which I.have

'but so that you would know the love that I have ...' 2 Cor 2:4
- (18) monon **tōn ptōchōn** hina mnēmoneuōmen
only the poor that we.may.remember —

'only that we should remember the poor' Gal 2:10
- (19) kai **tēn ek Laodikeias** hina kai hymeis anagnōte
and the.(one) from Laodicea that also you you.may.read —

'and that you read the one from Laodicea' Col 4:16

Once we have a clear understanding of this restriction on extracted constituents in Greek we are able to look at verses such as 1 Cor 14:12 and 1 John 5:16 from a better vantage point. These are the only cases in the New Testament (to my knowledge) which might be taken as counter-examples to the proposed constraint on extraction. I discuss 1 Cor 14:12 first. The phrase in question is in boldface.

- (20) epei zēlōtai este pneumatōn pros tēn oikodomēn
 since zealots you.are of.spiritual.(things) to the edification
- tēs ekklēσίας zēteite hina perisseuēte
 of.the church seek (so?).that you.may.abound

A common understanding of this verse takes the prepositional phrase **pros tēn oikodomēn tēs ekklēσίας** 'to the edification of the church' as a constituent (directly or indirectly) of the **hina** clause. The NIV is representative: "Since you are eager to have spiritual gifts, try to excel in gifts that build up the church". Under this analysis, however, the extracted phrase occurs in a position which is not licensed by the grammar of Greek.⁶ As a constituent of the embedded clause, it should appear after **zēteite** 'seek', not before it.

There is an alternative analysis which is consistent with the observed constraint on extraction and the context. This is for the **pros** 'to' phrase to be taken as a modifier of the **zēteite** 'seek' clause. If the **hina** clause is then taken as the complement of **zēteite**, a possible rendering is: "Since you are eager to have spiritual gifts, for the building up of the church try to have more (of them)."⁷ If the **hina** clause is taken as a purpose clause, a possible rendering is: "Since you are eager to have spiritual gifts, seek them [or: strive] for the building up of the church, in order that you may have more (of them)".⁸

Since these alternative renderings are as viable contextually as the position adopted by the NIV (*inter alia*), there is no warrant from this verse to abandon the proposed constraint on extraction.

The next possible counterexample is more difficult since the common rendering is without much dispute. In fact most commentators do not mention any difficulty with the syntax.

- (21) estin hamartia pros thanaton
 (there).is sin to death
- ou peri ekeinōs legō hina erōtēsē
 not concerning that I.say that he.may.ask

'There is a sin that leads to death. I am not saying that he should pray about that.' 1 John 5:16 (NIV)

One question is whether the negative **ou** is to be construed as negating the verb **legō** 'I say' (as many translations render it) or as negating the phrase **peri ekeinōs** 'concerning that'. Under the first view, there is little room for doubt that the prepositional phrase has been extracted from the complement clause and placed in an unusual position--between the negative word and the higher verb. This would be a strong counterexample to the proposed constraint on extraction. Under the second view, it is possible to take the construction as a kind of cleft construction, rendered as follows in the Louis Segond French translation:

- (22) Il y a un péché qui mène à la mort;
ce n'est pas pour ce péché-là que je dis de prier.

'There is a sin that leads to death; it is not for that sin that
I say to pray'

This understanding certainly fits well with the context. If this is indeed correct, then this verse does not involve extraction, but rather clefting. Cleft constructions have long been recognized as having much in common with WH-Movement; therefore it is not surprising to find the prepositional phrase where it is.

In conclusion, we have seen that there appears to be a significant restriction on the extraction of elements from a finite embedded clause. Given this constraint we are forced to pay more attention to less commonly suggested renderings of verses which appear to counterexemplify it.

Notes

* I thank the following people for their comments on various drafts of this paper: Bruce Hollenbach, Bruce Turnbull, Desmond Derbyshire, Robert Smith, and Richard Young. Of course, none of them is responsible for the shortcomings that persist despite their help. Except where noted as being taken from the New International Version (NIV), the translations of the verses are fairly literal personal translations of the 21st edition of the Nestle text, punctuation omitted.

1. Robertson 1934:423 indicates that he is not unaware of the facts discussed here. But he lumps various phenomena under the rubric "prolepsis," including extraction and raising (see below). The same is true of Turner 1963:325. Robertson also limits prolepsis to "substantives," which is not correct for extraction, as can be seen from the examples presented below. Winer 1881:625-9 discusses the subject very clearly under the name 'attraction', but also includes other constructions with it.

2. See, for example, Radford 1981.

3. See Marlett 1976, Joseph 1978, and Marlett 1984. Actually, word order facts suggest that infinitival complements in Greek result not from raising, but from clause union.

4. See Marlett 1976 and Willson 1985. Marlett 1976 discusses this construction under the name left dislocation, while Willson uses the name extraposition; these names are perhaps best reserved for other constructions. The term 'extraction' is a bit too general in that various kinds of movements, including WH-Movement, are also generally considered extractions. Nevertheless, for lack of a better name I will use the name extraction in this paper.

5. I do not discuss what happens when the subordinate clause is non-

finite except to note that the constraints are different, as Willson 1985 points out. (See note 3.) Also, it should be noted that the TEV does not take Rom 11:31 as a case of extraction; and various commentaries, including Robertson 1934:425, Blass and Debrunner 1961:291, and Morris 1958, do not take 1 Cor 15:2 as a case of extraction. The argument by Morris against the extraction analysis is without force, however, since one cannot simply dismiss an extraction analysis as "not a very natural Greek construction" (204). Other possible examples of extraction in the NT include Luke 7:39, Gal 5:11, 1 Cor 14:10, and Acts 19:2.

6. Godet 1977:704 cites 1 Cor 3:5, 7:17, and 9:15 as being other examples of the "inversion" of word order seen in 15:2. His argument is flawed, however, since in these other cases clause boundaries are not involved.

7. Three commentaries which take a position along these lines are Robertson and Plummer 1914, Meyer 1884, and Barrett 1968. I am not taking a position here on the proper translation of the verb *perisseuōte*.

8. Alford 1968:593, Findlay 1974:905, and Conzelmann 1975:237 view the construction in this way.

References

- Alford, H. 1968. The Greek Testament, vol. 2. Chicago, Moody.
- Barrett, C.K. 1968. A commentary on the first epistle to the Corinthians. New York: Harper & Row.
- Blass, F. and A. Debrunner. 1961. A Greek grammar of the New Testament and other early Christian literature, edited by R. Funk. Chicago: University of Chicago.
- Conzelmann, H. 1975. 1 Corinthians. Philadelphia: Fortress.
- Findlay, G. 1974. St. Paul's first epistle to the Corinthians, The expositor's Greek Testament, vol. 2. Grand Rapids: Eerdmans.
- Godet, F. 1977. Commentary on First Corinthians. Grand Rapids: Kregel.
- Holy Bible, The: New International Version. 1978. Grand Rapids: Zondervan.
- Interlinear Greek-English New Testament, The: The Nestle Greek text with a literal English translation. 1959. 2d ed. Translation by A. Marshall. Grand Rapids: Zondervan.
- Joseph, B. 1978. Morphology and universals in syntactic change: evidence from medieval and modern Greek. Ph.D. Thesis, Harvard University.

- La Sainte Bible qui comprend l'Ancien et le Nouveau Testament traduits d'après les textes originaux Hébreu et Grec. 1959. Translated by Louis Segond. British and Foreign Bible Society.
- Marlett, S. 1976. Copy-raising in Koine Greek. M.A. Thesis, University of North Dakota.
- Marlett, S. 1984. "Switch-Reference and subject raising in Seri", in E. Cook and D. Gerdts, editors, The syntax of native American languages. New York: Academic.
- Meyer, H. 1884. Critical and exegetical handbook to the epistles to the Corinthians. New York: Funk and Wagnalls.
- Morris, L. 1958. The first epistle of Paul to the Corinthians: an introduction and commentary. Grand Rapids: Eerdmans.
- Radford, A. 1981. Transformational syntax: a student's guide to Chomsky's extended standard theory. Cambridge: Cambridge University.
- Robertson, A. 1934. A grammar of the Greek New Testament in the light of historical research. Nashville: Broadman.
- Robertson, A. and A. Plummer. 1914. A critical and exegetical commentary on the first epistle of St. Paul to the Corinthians. Edinburgh: T. & T. Clark.
- Turner, N. 1963. Syntax. Edinburgh: T. & T. Clark.
- Willson, S. 1985. Extraposition in New Testament Greek, ms., University of North Dakota.
- Winer, G. 1881. A grammar of the idiom of the New Testament, trans. E. Masson, ed. and rev. J. Thayer, 7th ed. Andover, Mass: Draper.

**FROM MEANING TO TEXT:
Semantic Representation in the Meaning-Text Linguistic
Theory and a New Type of Monolingual Dictionary**

Igor A. Mel'čuk
University of Montreal

- 1 Introduction
- 2 Outline of the Meaning-Text Theory
 - 2.1 Main postulates and general characteristics
 - 2.2 Semantic representation in the Meaning-Text Theory
- 3 The Explanatory Combinatorial Dictionary as a crucial component of a Meaning-Text model
 - 3.1 General characteristics
 - 3.2 Structure of an ECD entry
 - 3.2.1 Semantic zone
 - 3.2.2 Syntactic zone
 - 3.2.3 Lexical cooccurrence zone
- Appendix A: List of Lexical Functions
- Appendix B: three sample dictionary entries

1 Introduction

We use language primarily to express meanings. And to communicate, of course, but "**sprachlich**"¹ communication boils down to an exchange of meanings expressed in and extracted from "sprachlich" messages). This is an obvious truth or, rather, a commonplace fact that no one in the linguistic profession has ever tried to call into question. All linguistic schools include in their respective credos strong statements concerning the crucial role of meaning and its expression in human language and, consequently, in linguistics. Thus half a century ago Leonard Bloomfield wrote: "In human speech, different sounds have different meanings. To study this coordination of certain sounds with certain meanings is to study language" (1933: 27). More recently, the same tenet has been vigorously restated, by (among others) Noam Chomsky: "A generative grammar ... is a system of rules that relate signals to semantic interpretations of these signals" (1966: 12). "The grammar of a language ... establishes a certain relation between sound and meaning" (1968: 116); etc. For the last ten years general interest in meaning and in the linguistic discipline dealing with meaning, i.e. semantics, has witnessed considerable growth.

Nevertheless, semantics today still remains an underdeveloped field. Meaning, which is **de jure** recognized as an important element of language, still is **de facto**, if not ignored, then frequently avoided or at least not dealt with as directly and systematically as it should be. Let us mention, for instance, Testen et al. 1984 - a volume called

Lexical Semantics and representing one of the most recent publications in the field. The volume contains many interesting papers, and is full of astute remarks and insightful discussions of different semantic problems, but, curiously enough, in spite of its title, it does not offer descriptions, or even attempts at descriptions, of actual lexical meanings. This is quite typical of the approach of modern linguistics to semantics: First, most linguists concentrate on abstract questions having a strong logical flavor (quantification, logical connectives, referentiality, pragmatic functions of discourse, etc.), while consistently shunning the description of genuine "sprachlich" meanings, lexical or grammatical. Second, to the best of my knowledge, there is, in the mainstream of modern linguistics, no integrated theory of language that would consistently proceed from meaning. There is, at least, one obvious exception: the work of A. Wierzbicka proposes, develops and substantiates such a theory (cf. below). To discuss the reasons for this state of affairs would be out of place; suffice it to state that it justifies our expounding here a different approach to natural language, namely, the Meaning-Text Theory.

The Meaning-Text Theory (MTT) is not simply another of numerous linguistic theories, which proliferate to such an extent nowadays that there are almost as many theories as there are practicing linguists. The MTT is truly different in that it puts quite a new emphasis on "sprachlich" meaning, taking it as a cornerstone of language description as a whole; accordingly, semantics is declared and, most importantly, actually treated as the central linguistic discipline (which, among other things, underlies syntax and morphology).

The Meaning-Text Theory is by no means a novelty. It was launched in 1965, in Moscow, by A. Zholkovsky (now at the University of Southern California) and the present author: Žolkovskij and Mel'čuk 1965. Later, we were joined by Ju. D. Apresjan. Since that time, the MTT has been developing over the past 20 years (see Mel'čuk 1981 and Nakhimovsky 1983). However, the MTT still lacks an easily available and sufficiently detailed presentation in English, a gap which this paper tries to fill, at least in part, by outlining the two following topics:

- (i) General structure of the MTT (with special attention to its semantic chapter) (Sect. 2).
- (ii) A new type of monolingual dictionary, or lexicon, which according to the MTT constitutes the central part of the semantic component of any full-fledged "sprachlich" model (Sect. 3).

The character of our exposition forces us to ban references, including the most relevant ones. (Only a few are quoted, where they were judged necessary in order to identify the background of the MTT.)

2 Outline of the Meaning-Text Theory

2.1 Main Postulates and General Characteristics

The basic idea underlying the Meaning-Text Theory is as follows. Any speech act is believed to consist of three major components:

- (a) A CONTENT to be communicated by "sprachlich" signals; it will be called **meaning**.
- (b) A complex SIGNAL, or SIGNALS (observable physical phenomena), used to communicate the content in question; this signal will be called **text**.
- (c) A CORRESPONDENCE between the meaning and the text; it will be called **mapping**.

We will shortly make the concepts of meaning and text more precise, but before we do so we need to emphasize the following three important points concerning our tripartite division of a speech act.

First, we presuppose the discrete character of both meanings and texts. That is, we postulate that we can distinguish meanings and count them; likewise for texts. Meanings and, *mutatis mutandis*, texts are for us formal objects specifiable by a logical device (for instance, by a generative grammar). This is more or less accepted with respect to texts but seems to be rather unusual as applied to meanings.

Second, the set of possible speech acts is open: there is an infinite number of meanings and texts in every natural language. However, the mapping between meanings and texts is finite, i.e., it can be fully represented by a finite number of correspondence rules (even though this number is very large). This is the case since an arbitrary meaning, as well as an arbitrary text, can be, generally speaking, broken down into simpler meanings, viz. simpler texts, so that ultimate correspondences can be established between elementary meanings and elementary texts, the number of these entities being finite.

Third, there is no one-to-one correspondence between meanings and texts: one meaning can be expressed by quite a few texts (synonymy), and a text can express several meanings (homonymy/polysemy — or ambiguity).

The above-stated may be symbolized as in (1), which is the main postulate of the Meaning-Text Theory:

- (1) **Postulate 1**
- $$\{\text{MEANING}_i\} \xrightarrow{\text{language}} \{\text{TEXT}_j\} \mid 0 < i, j < \infty$$

In the MTT, then, a natural language is viewed as a many-to-many mapping between an infinite set of meanings and an infinite set of texts.

Let it be emphasized and re-emphasized that the words **meaning** and **text** are to be taken, within the present framework, as technical terms, free of the many connotations they have in every-day English or in other terminological systems. Thus, as far as meaning is concerned, three important properties of our approach must be constantly borne in mind:

(i) We deal only with strictly "sprachlich" meaning, i.e. with the information which can be extracted from an utterance solely on the basis of purely "sprachlich" knowledge, without any reference to the context and/or extralinguistic environment (roughly, without any encyclopedic knowledge about the universe). Our "meaning" is the shallowest, absolutely literal meaning of utterances. Suppose a girl on a date says to her shy partner, "Oh, I'm cold!", in order to incite him to embrace her; for the MTT, the meaning of her utterance is still 'I'm cold' and by no means 'Go ahead and embrace me!' When I tell you that John has applied for a scholarship and you react by asking me, "What do you mean?", you are after quite a different meaning from what the MTT tries to describe. A meaning in the MTT is simply the invariant of a set of all utterances that are paraphrases of each other. Or, to put it differently, **meaning is the canonical invariant of synonymous utterances** (see below). Therefore, such things as truth, incoherence or absurdity are of no relevance to us: these properties concern the relationship between "sprachlich" meaning and something else (= the universe), while the task of the MTT is precisely to avoid discussing anything that is beyond language.

(ii) Meaning is taken to be directly accessible to speakers — much like the sounds of their language. This is not to be construed as implying that any speaker is able to correctly analyze or explain any given meaning; such is by no means the case. But every speaker knows (though perhaps subconsciously) what he wants to say, even if he does not understand the deeper meaning of his words (this last circumstance should not bother us). Therefore meaning is open to direct introspection; for a trained linguist describing his mother tongue, meaning belongs to the data.

(iii) Although we talk of meaning, the MTT deals, in actual fact, with formal representations of meanings, called Sem(antic) R(epresentations). Meanings do have an objective existence of their own — as certain neurophysiological events in the speaker's brain. But once again, this is of no relevance for us. When we write {MEANING₁}, we mean {SemR₁}. These representations, invented and developed by the linguist, are our only semantic reality within the limits of the MTT, of course.

Texts possess three similar properties as well.

(i') The term **text** should by no means be understood as 'coherent text' or 'discourse', i.e. in the sense of the so-called text grammar. Our text is any linguistically₁ valid segment of speech; it can be the signifier (Lat. **signans**, Fr. **signifiant**) of a morph, a lexeme, a phrase, etc., including whole volumes.

(ii') Texts are directly accessible to the speaker, in the sense that he is able to tell a well-formed text in his language from an ill-formed or questionable one.

(iii') We deal with representations of texts only, not with actual texts (which consist of physical phenomena). Saying {**TEXT**_j}, we mean {Phon(etic) R(epresentation)_j}, and only PhonR's are the object of our study.

Following the statements (i)-(iii) and (i')-(iii'), we can rewrite (1) as a more precise formulation, namely (2):

$$(2) \quad \begin{array}{c} \text{language} \\ \{\text{SemR}_i\} \quad \langle \text{---} \rangle \quad \{\text{PhonR}_j\} \mid 0 < i, j < \infty \end{array}$$

What has been said so far allows us to draw an important conclusion. Given that we have to deal with representations of meanings and texts, our first and foremost task should be to develop formal languages to represent them. In slightly different words, the first thing we need in order to talk about natural language in a scientific way, so that all our statements are explicit and unambiguous, is a semantic transcription and a phonetic/phonemic transcription. Until we can formally present meanings and texts of a natural language we cannot, strictly speaking, discuss it.

Now, phonetic/phonemic transcriptions are readily available; where they are not, methods for developing them are familiar, and necessary skills and know-how are at hand. With meanings, however, the situation is different -- quite simply, it is disastrous. No serious efforts have been made in mainstream linguistics to develop semantic transcriptions; not even one artificial semantic (meta)language for one natural language has been proposed, as far as we know, in major linguistic schools. There are, to be sure, several developments that seem to contradict the last statement: take, for example, the insightful and promising work of A. Wierzbicka (1972 and 1980, to name only two of her many pioneering studies). But in spite of its interest and availability, Wierzbicka's research (as well as a number of similar attempts) remains little-known and, in a sense, marginal. Modern linguistics, as a whole, has not yet made semantic language its main concern. In sharp contrast, the MTT lays heavy emphasis on meaning representation, i.e. on SemR and, therefore, on a semantic language. In our opinion, such a language is a *sine qua non* of today's science of language. Later on, we will discuss

the semantic language, i.e. the language of SemR used in the MTT, in more detail.

Now let us make the next important point. While meanings and texts (of a language) are immediately accessible to the linguist, rules that correlate them (e.g. the mapping "<==>" in (1) and (2)) are not. Thus we linguists face the classical "black box" situation: we are in front of a running device (natural language) whose inputs and outputs are perceptible and controllable but whose inner structure, the circuits linking inputs with outputs, is totally hidden from us. "Sprachlich" rules mapping meanings onto texts and vice versa are electronically/chemically encoded in the speaker's brain, yet we cannot open skulls to find out about them. The only option we have can be expressed by the following:

(3) **Postulate 2**

A natural language, viewed as a many-to-many mapping between two infinite sets (that of meanings and that of texts), can be described only by a **Cybernetical, or Functional, Model**.

A **functional model** of a device (phenomenon, event) X is a logical device, or a finite set of formal rules, which simulates as closely as possible the behavior of X. In our case, X is human language, and therefore our model should simulate the "sprachlich" behavior of humans -- in other words, it should establish correspondences between meanings and texts. It is such models that we call **Meaning-Text Models (MTM)**. Strictly speaking, an MTM for language $\mathcal{L}$ is a system of formal rules which associates with a given SemR of $\mathcal{L}$, all the PhonR's that, in the judgment of $\mathcal{L}$'s speakers, can carry the corresponding meaning, and vice versa, which associates with a given PhonR of $\mathcal{L}$ all the possible SemR's -- all meanings that this text can have. Thus our second task should be to develop Meaning-Text models of languages. (Of course, this task is closely related to the first one, that of developing formal languages for linguistic representations in the first place -- for the SemR.) We can say that the Meaning-Text Theory is a theory for building MTM's.

Note that a complete Meaning-Text model should be a "dynamic" device in the sense that for a given SemR, it should actually produce the set of corresponding PhonR's. However, it seems natural to distinguish, within an MTM, two submodels: 1) the system of purely "sprachlich" rules which specify the correspondence between meanings and texts, and 2) the system of procedural rules which specify the process leading from a SemR to the PhonR's (or vice versa) -- based on "sprachlich" rules, i.e. on the first system. The second system is by no means specific to linguistics. It includes rules that, using factual knowledge supplied by the first system, compute the best way (or one of the best possible ways) to construct the necessary PhonR for a given SemR (and to construct the necessary SemR for a given PhonR); the same type of procedural rules are needed whenever stored knowledge has to be

used for the solution of any problem. Therefore, in all our deliberations, we will not concern ourselves with the second system of rules. Thus an MTM is considered and presented only in its strictly "sprachlich"/linguistic "static" aspects. This means, among other things, that it appears as a **Static** system of correspondences between elementary meanings and elementary texts, to the complete exclusion of all elements of procedure. Two important remarks are needed in this regard.

First, an MTM is by no means a generative or, for that matter, transformational device: it is **equative** (or, if you like, translatable). It does not seek to generate (enumerate, specify) the set of all (and only) grammatical or meaningful texts, nor does it transform certain "sprachlich" entities into other entities. It simply states which SemR's match which PhonR's in accordance with native speakers' intuition.

Second, an MTM does not admit linear order as a formal means among its tools. To be sure, linear order is an important expressive means of natural languages and, as such, it has to be accounted for in the rules and the representations used by the model. However, "sprachlich" order should not be reflected simply by ordering the elements of the MTM. Thus MTM rules should be unordered and all relevant information about language should be explicitly represented by symbols and configurations thereof. Likewise, linear order is never admitted to express any "sprachlich" relations in the semantic or syntactic representations of utterances. Once again, all relations should be denoted by appropriate symbols. (Cf. Sanders 1972.)

We have mentioned above that the correspondence between meanings and texts is many-to-many. For a fairly complex meaning, hundreds of thousands of (nearly) synonymous texts can be constructed. Thus, example (4)

- (4) The Food and Drug Administration has seriously
 cautioned expectant mothers to avoid one of
 life's simple pleasures: a cup of coffee.

has more than 200,000 paraphrases (Mel'čuk 1981: 31-32). Similarly, a text may have several meanings (i.e. have distinct readings). The extremely involved character of the Meaning-Text correspondence makes it practically impossible to write its rules directly from meanings to texts. It turns out that for a better perspicuity and surveyability, the correspondence between meanings and texts must be broken down into simpler components. This entails the following:

(5) **Postulate 3**

To describe the correspondence $\{\text{SemR}_i\} \langle == \rangle \{\text{PhonR}_j\}$,
TWO INTERMEDIATE LEVELS of utterance representation
are introduced: Synt(actic) R(epresentation) and

Morph(ological) R(epresentation).

Therefore, (2) can now be rewritten, in a more developed form, as (6):

(6)

$$\underbrace{\{\text{SemR}_i\}}_{\text{Semantics}} \langle === \rangle \underbrace{\{\text{SyntR}_k\}}_{\text{Syntax}} \langle === \rangle \underbrace{\{\text{MorphR}_l\}}_{\text{Morphology}} \langle === \rangle \underbrace{\{\text{PhonR}_j\}}_{\substack{\text{+} \\ \text{Phonology}}}$$

The SyntR and MorphR are centered, respectively, around the sentence and the word. Because of the obvious space constraints, we will not touch here upon such important properties of the SyntR and the MorphR as their division into deep and surface sublevels, the formalisms used, etc. We will limit ourselves to insisting on the stratificational, or multistratal, character of the proposed model (cf. Lamb 1966 and Sgall 1967). A Meaning-Text "sprachlich"/linguistic model consists, as shown in (6), of three major components, which form, so to speak, a production chain:

Semantics establishes the correspondence between the infinite set of SemR's and the infinite set of SyntR's;

Syntax establishes the correspondence between the SyntR's and the MorphR's;

Morphology + Phonology establish the correspondence between the MorphR's and the PhonR's.

In this presentation, we will be concerned exclusively with the semantic component of an MTM.

Note that in principle the correspondence "<===>" is bidirectional: an MTM should be able to go both from meanings to texts (= speech production) and from texts to meanings (= speech understanding). Logically, both directions are, of course, equivalent, and both correspond to actually observable processes. Linguistically, however, they are not equivalent; natural language gives a more prominent place to the speaker than to the addressee:

- (i) The speaker can speak (or write) even without an explicit addressee (to himself, to God, to posterity), while an addressee is inconceivable without an explicit speaker.
- (ii) Reference to, or value judgments by, the speaker (rather than by the addressee) are included in the meaning of a hčst of "sprachlich" units: shifters (like **here** = 'where the speaker is'), speaker-oriented verbs like **come** vs.

go, performative verbs, all affective and derogatory terms, and many others, without mentioning the cases of empathy (the speaker identifying himself with a character presented in his utterance), prepositional choices (**X in front of Y** as opposed to **Y behind X**, as a function of where the speaker mentally places himself), etc.

- (iii) Most important, the speaker knows perfectly well what he is going to say; his only problem, when he starts from a given meaning, is to find an appropriate "sprachlich" form (= text) to convey this meaning. His activity *qua* speaker is a purely "sprachlich" one. The addressee, on the other hand, has no previous knowledge of the meaning he is supposed to extract from the utterance. He has to use his logic, his general abilities to infer and guess, his knowledge of the world, etc. — in addition to his "sprachlich" skills proper. The addressee's activity *qua* addressee is not purely "sprachlich": it includes, to a considerable degree, many non-"sprachlich" operations, with the result that "sprachlich" operations occupy in it a rather modest place.²

Consequently, the viewpoint of the speaker is by far the more advantageous for linguists. Describing language as a system for EXPRESSING meanings and not the other way around, we can concentrate on genuine "sprachlich" phenomena, thus avoiding involvement with problems of rule ordering and application, modalities of processing, encyclopedic understanding, and the like, which are absolutely irrelevant to linguistics. Let it be emphasized that certain "sprachlich" phenomena can be properly recognized and systematically studied only if considered in the direction from meanings to texts. Thus the following oppositions do not represent any interesting problem from the viewpoint of the addressee:

- (7) a. **strong** warning vs. **considerable** attention
 <*considerable warning, *strong attention>;
- b. They differ **widely** vs. They miss her **acutely**
 <*differ acutely, *miss widely>;
- c. **much/widely/greatly** publicized vs. **heavily** settled [area] <*heavily publicized, *much/widely/greatly settled>.

A listener or a reader easily understands the expressions in (7) and misses the interesting fact that the boldfaced lexemes express one and the same meaning, namely 'very', and are (complementarily) distributed according to the lexeme modified. In fact, these lexemes are values of

what is called the lexical function **Magn**. (The term **Magn** is from Lat. **magnus** 'great, big'.)

- (8) **Magn**(warning) = strong **Magn**(attention) = considerable
 Magn(differ) = widely **Magn**(miss) = greatly
 Magn(publicized) = much, **Magn**(settled) = heavily
 widely,
 greatly

To discover that natural languages feature a limited number of very abstract and general meanings such that any one of these has numerous expressions distributed as shown above, one needs to look at language from the speaker's viewpoint. (Cf. the discussion about lexical functions, Sect. 3.2.3 and a list of lexical functions, Appendix A.)

Therefore, in sharp contrast with most linguistic schools (traditional grammar, European and American structuralism, transformational grammar and its various outgrowths), the Meaning-Text Theory insists on the meaning-to-text direction in linguistic research, as well as in linguistic description.

To sum up: The Meaning-Text Theory aims at developing Meaning-Text models of natural languages. An MTM is taken to be a system of rules which establish correspondences between semantic and phonetic representations of utterances, the SemR's being written in a special semantic language, devised by the researcher. This is a multilevel, or multicomponent, system, with a semantic component responsible for the correspondence between semantic and syntactic representations: the semantic component "translates" SemR's into SyntR's (and vice versa). Last, but not least, the direction of research and description is that of production: from meanings to texts.

2.2 Semantic Representation in the Meaning-Text Theory

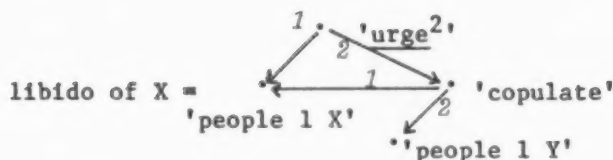
In conformity with the goal of this paper, we will now try to throw light on the concept of semantic representation.

Let us begin with an example. In Toronto's **The Globe and Mail** (Aug. 5, 1985, N6) we find the following title:

- (9) The death of libido.

This is a short editorial, claiming that watching situation comedies on American TV has a pernicious effect on the viewer's sexual drive. The meaning of this title is clear; how can we represent it? Well, **libido** means (roughly) 'urge² to copulate' (superscripts and numbers after words in our semantic descriptions refer to intended senses in **Longman Dictionary of Contemporary English**). 'Urge²' [noun] is by necessity someone's urge; in this context, an indefinite group of people is meant: 'people 1'. Then, taking 'urge²' and 'copulate' to be two-place predicates³, we can write:

(10)

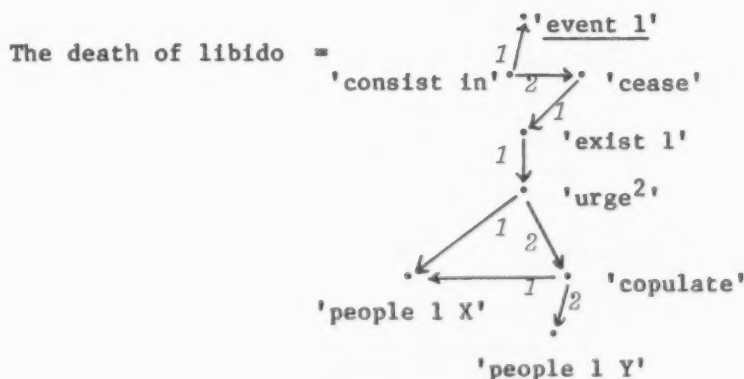


Or, to express it in prose,

(10') libido of X = 'urge² of people 1 X to copulate
with people 1 Y'

The noun 'death' is used in (9) metaphorically, and means 'event 1 consisting in that X ceases to exist 1'. In our formalism, then, we have (11):

(11)



Numbers assigned to arrows in diagrams of (10) and (11) identify different arguments of the same predicate, underlining singles out the main component.

We believe that (11) represents the "sprachlich" meaning of the phrase **the death of libido** quite well. Two qualifications seem, however, to be in order.

First, while the expressions in single quotes in (10) and (11) look like English words, they are NOT English words -- rather, they are handy designations of specific senses of the corresponding English words, which, like almost all English words, are polysemous. To disambiguate them, we are using sense numbers borrowed from the **Longman Dictionary**, as indicated above. (For our illustrative purposes, it could be any other dictionary.) A word taken in one well-specified sense is called a **lexeme**. Then, the expressions in single quotes are **semantic units**, or **sememes**, notated with English lexemes. These sememes do not feature syntactic and morphological properties typical of English lexemes that represent them. They are indeed units of meaning.

Second, one might ask the legitimate question about the meaning of

sememes themselves. For example, what does **cease** mean? In this case, the answer is simple:

$$(12) \quad \text{cease to P} = \begin{array}{c} \text{'begin 1' 'not' 'P'} \\ \xrightarrow{1} \xrightarrow{1} \end{array}$$

[John ceased to work = 'John began not to work']

But then, what is "begin 1"? The answer is still at hand:

$$(13) \quad \begin{array}{l} \text{begin 1 to P} = \text{'at moment } t_0, \text{ P does not exist 1;} \\ \quad \quad \quad \text{at moment } t_1, \text{ P exists 1;} \\ \quad \quad \quad t_1 > t_0 \end{array}$$

However, things do not always go as smoothly as this. What does **exist 1** mean? And **moment 1**? And **not**? For the time being, I cannot answer.

I do not imply that these terms cannot be explained or that their meaning cannot be described at all. It can, but maybe NOT IN "SPRACHLICH" TERMS. From the viewpoint of strictly "sprachlich" meaning, these sememes could be elementary or indecomposable: they are then **semantic primitives**. (Cf. 13 semantic primitives postulated by A. Wierzbicka: 1972 and 1980.⁴)

Now we can proceed to specify the formal language, or **semantic metalanguage**, in which semantic representations within the MTT framework can be written. (To simplify the presentation, we will not consider the communicative organization of meaning, i.e. contrasts of topic vs. comment, given vs. new, etc.)

Like any other formal language, the semantic language of the MTT is a pair of sets: a finite set of elements called **alphabet** (or **lexicon**) and a finite set of expressions called **rules**; there are **formation rules**, which specify well-formed configurations of elements, and **transformation rules**, which specify equivalences between well-formed configurations. Let us take these components of the semantic language in turn:

An expression in the semantic language, i.e. a Sem(antic) R(epresentation), is, formally speaking, a **completely labeled network**: a connected oriented graph whose **vertices**, or **nodes**, are labeled with sememes and whose **arcs**, or **branches**, are labeled with arbitrary distinctive symbols. (We will use Arabic numbers; the purpose of using them is, as indicated above, to distinguish different arguments of the same predicate, cf. below).⁵

The **alphabet** of the semantic language in question includes:

(a) A tiny list of formal elements (used to construct networks): vertices, represented by points; arcs, represented by arrows; and a

small set of natural numbers as distinguishers, appearing as labels on arcs.

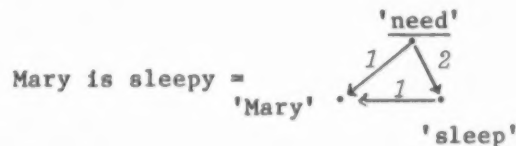
(b) A huge list of sememes (several million, generally speaking): each sememe corresponds to a specific sense of a specific word of the language in question. A sememe can be elementary, and then it is called a **seme**; otherwise, it is complex and is representable in terms of other sememes and/or semes.

Sememes fall into two major classes:

1. **functors**, or sememes having "places" for other sememes, which are then their **arguments**, and
2. **names** (of (classes of) objects), which may only serve as arguments but may not have their own arguments.

The relationship between a functor and its i -th argument is shown, in the semantic language, by an arrow labeled i and pointing from the functor to the argument. For example, 'sleep' is a one-place functor, and therefore, **Mary sleeps** = 'Mary' 'sleep'; 'need' is a two-place functor, so that

(14)



Functors are meanings corresponding to actions (such as 'unite', 'kiss'), states ('sleep', 'disappear', 'joy'), properties ('lovely', 'blue'), relations ('more than', 'be part of'), parameters ('the speed of', 'the price of'), events ('rain', 'explosion', 'fall'), quantities ('many', 'three thousand'), etc. Names are meanings corresponding to objects, substances, natural species, etc.

Functors are subdivided, first according to the number of the arguments they can take (one-place, two-place, etc.; in actual fact, natural languages do not show functors with more than 6 arguments); and second, according to the restrictions on the semantic type of the arguments they can take (both arguments of 'cut' must be names; the only argument of 'begin' must be a functor; the first argument of 'see' must be a name, but the second may be either a name or a functor). There are three major types of functors: predicates, quantifiers and logical connectors, but we will not delve into this matter.

Formation rules for SemR's are trivial and largely obvious. The following are some of the more basic ones:

- a SemR must be a connected, oriented, labeled graph

- the number of arrows leaving a node which is labeled with a functor must correspond to the number of places of this functor
- no arrow should leave a node labeled with a name
- all the restrictions inherent in functors which appear in a SemR should be observed

Transformation rules for SemR's are equations which correlate (non-elementary) sememes with their **decompositions**: 'die 2' = 'cease to exist 1' = 'begin 1 to not exist 1', etc. Note that, strictly speaking, we should use semantic networks and write as follows:

$$(15) \quad \begin{array}{ccc} \begin{array}{c} \downarrow 1 \\ \text{'die 2'} \\ \downarrow \\ \text{'X'} \end{array} & \Longleftrightarrow & \begin{array}{c} \downarrow 1 \\ \text{'cease'} \\ \downarrow \\ \text{'exist 1'} \\ \downarrow 1 \\ \text{'X'} \end{array} \end{array}$$

However, here as everywhere below, we allow ourselves to simplify the presentation, if this does not entail ambiguities or misunderstandings.

By means of such rules we can carry out equivalent transformations of SemR's:

'His desire 1 died 2' = 'His desire 1 ceased to exist 1' = 'His desire 1 began 1 to not-exist 1'.

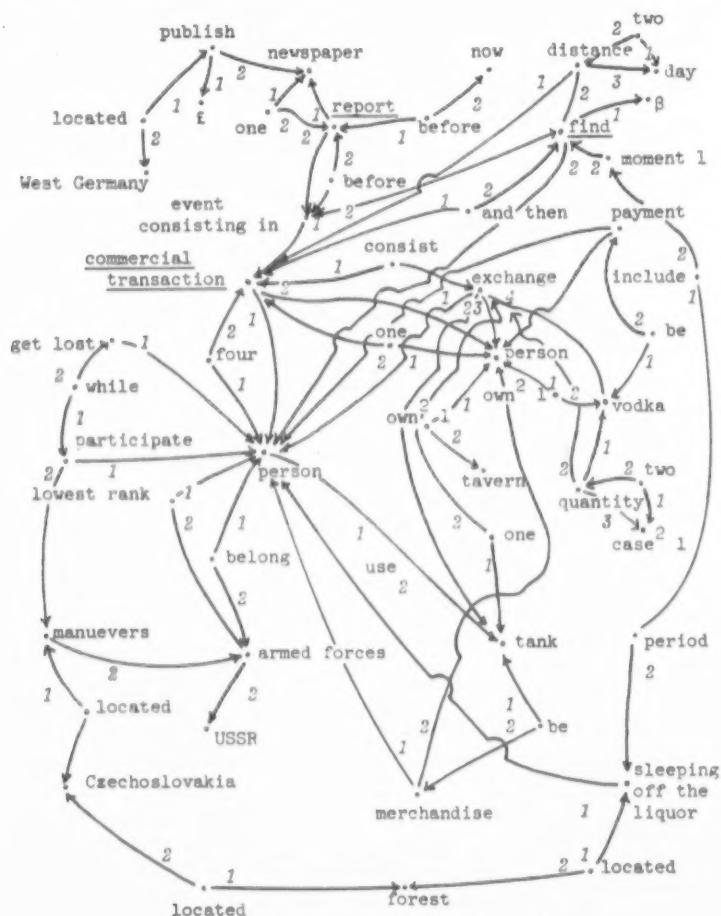
Let it be emphasized that a rule of the form 'X dies 2' = 'X ceases to exist 1' is nothing more than a lexicographic definition of a particular lexical sense of an English word. It is here that the concept of a special dictionary enters the scene; we will return to it shortly.

Put into a nutshell, the semantic language for, say, English, is a hybrid using the grammar of oriented labeled networks and the English vocabulary, but in a disambiguated form. We take as basic lexical units of the semantic language not words, but word senses defined in terms of more elementary senses, which in their turn are defined in terms of still more elementary senses, and so on, until we arrive at semantic primitives; these latter have to be specified by a list. To achieve this, our definitions should obey precise and rigorous rules (for instance, vicious circles must be banned), which we cannot discuss here. (A few words, though, will be said about them in Sect. 2.) Instead, let us consider a sample SemR: the SemR for a sentence taken from the same issue of **The Globe and Mail** as mentioned above (Aug. 5, 1985; page N6):

- (16) Four Soviet soldiers lost on maneuvers in Czechoslovakia traded their tank to a tavern owner for two cases of vodka and were found sleeping off the liquor in a forest two days later, a West German newspaper has reported.

We have to simplify our SemR drastically in order to make it surveyable. Thus, we do not even try to decompose many sememes which are readily decomposable. For instance, we used in the SemR, without further ado, the expressions **get lost** [= 'come to not know where is the way which one should follow'], **sleep off the liquor**, **maneuvers**, **exchange**, and others. The case of exchange would be especially interesting. Decomposing **X exchanges Y with Z for W** into 'X causes that Z owns Y, which previously X owned, having the goal of thereby causing -- and X thereby causes -- that Z causes that X owns W, which previously Y owned' would much better show the semantic link between soldiers, vodka and their sleeping off the liquor. Furthermore, sense-distinguishing numbers are not shown with the sememes. We cannot explain, either, many important details of SemR (17); we will, however, comment on four technical points.

- (17)



(i) All temporal relations, grammatical or otherwise, are represented in the same homogeneous way: by means of the sememe 'before'. Thus '(to) report before now' renders (very roughly!) the present perfect of 'has reported'; **X two days later than Y** appears as 'Y before X (and) the distance between X and Y is two days'; for **X were found sleeping...** we have 'someone found X, the moment of finding being included into the period of X's sleeping'. Similarly, all locative relations are shown in a standard way as well: the sememe '(to be) located in'.

(ii) The sememe 'use' in 'soldiers use tank' is applied very much like the expressions 'use the bathroom' <the telephone>, i.e. in the sense 'to use X the way X is designed to be used'; this is the meaning of the lexical function **Real** (see below).

(iii) To show the scope of the newspaper's report, a node labeled 'event consisting in ...' is used: the arrows leaving this node identify the two events ('(performed) a commercial transaction' and '(somebody) found...') that enter into this scope.

(iv) The sememes underlying semantic assertions (as opposed to presuppositions and modifications) are underscored, cf. above.

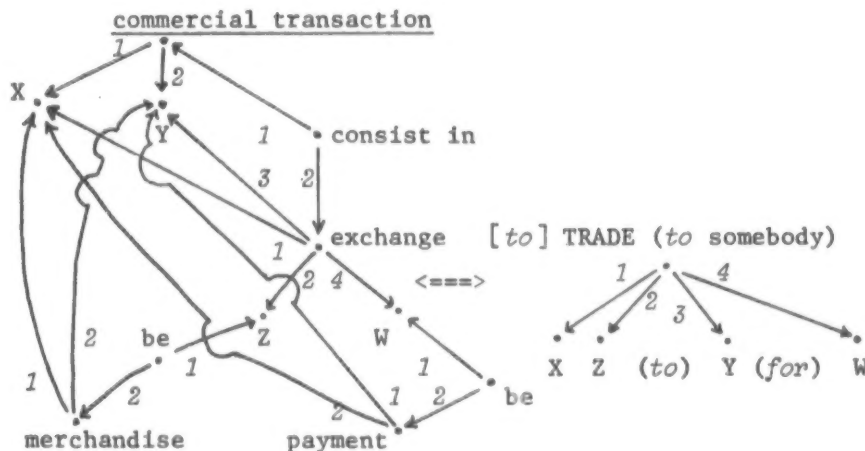
Strictly speaking, the network which appears in (17) is not a SemR; it is only a part of a SemR, namely the so-called Sem(antic) S(tructure) of sentence (16). A complete SemR must include two other parts: the Comm(unicative) S(tructure), which specifies the division of the given meaning into theme (topic) and rheme (comment), given vs. new information, and the like. But as indicated above, for simplicity's sake we suppress the CommS and the Rhet(orical) S(tructure), which specifies the "artistic intentions of the speaker, and, by obvious metonymy, allow ourselves to call (17) a SemR.

Notice that for pedagogical reasons, it might be more advisable to choose a simpler SemR (i.e. to present a shorter and more trivial sentence), as is often done to illustrate a point. We have, however, preferred not to amuse our readers with toys but rather to show them the real thing, even at the risk of frightening some of them away. The crucial feature of the MTT is that it does face the enormous complexity of natural language semantics, and we feel it is important for our readers to be fully aware of the extent of this complexity.

The SemR as proposed has a property which is central to the present paper: the elements of a SemR are, generally speaking, "smaller" (i.e. semantically simpler), than the lexemes which must be used in the corresponding sentence. Thus, 'person belonging to the armed forces of X and having the lowest rank therein' is 'soldier'; 'X (performs) a commercial transaction with Y consisting in exchanging by X with Y a Z as a merchandise for a W as a payment' is 'X trades Z to Y for W'; etc. In the Synt(actic) R(epresentation), however, the elements are of course actual lexemes. Therefore, the main bulk of semantic rules of a

language, that is, of the rules that correlate the SemR and the SyntR conveying the same meaning, is composed of rules having the following form:

(18)



There are other types of semantic rules as well; e.g., rules that map sememes onto morphological categories, such as tenses ('before now' <==> past tense), grammatical numbers, etc., which appear in syntactic structures. But semantic rules of type (18) are by far the most numerous: for a given language, they number between 10^5 and 10^7 — since every such rule represents a particular lexeme, or a word sense.

Now, a semantic rule of type (18) is nothing else but (with all additional data concerning cooccurrence) a DICTIONARY (= LEXICAL) ENTRY. From this it follows that semantics is, roughly speaking, a dictionary, but a very specific dictionary. We will characterize it in Sect. 3.

3 The Explanatory Combinatorial Dictionary as a Crucial Component of a Meaning-Text Model

3.1 General Characteristics

The Explanatory Combinatorial Dictionary (ECD) differs from more familiar monolingual dictionaries in that it is not a commercial dictionary which aims at a particular public, is adapted to fulfill a particular task, and is bound to observe various pedagogical, typographical and financial constraints. An ECD is conceived and developed as one part of a scientific (theoretical) description of the language under analysis and is thus a theoretical lexicon (cf. Lakoff 1973: 162-164). An ECD is to a "normal" dictionary what a theoretical grammar is to a language textbook. This fundamental property entails the following five features typical of an ECD:

1. An ECD is a **production-oriented** dictionary: it is intended to supply all the information which is conveyed by lexical units and which is necessary to express a given thought, and is oriented in conformity with the orientation of the meaning-text model.
2. An ECD is a **semantics-based** dictionary (that is why it is called explanatory). The definition of the entry lexeme, written in the special semantic language discussed above, serves as a basis for the description of the paradigmatic and syntagmatic relations of this lexeme.
3. An ECD is a **combinatorial** dictionary: it describes the syntactic and lexical collocations of the entry lexeme in the greatest possible detail.
4. An ECD is a **systematic** dictionary: it lays heavy emphasis on the homogeneity of the lexicographic descriptions. It is developed by lexical fields rather than by alphabet: similar lexemes have similar descriptions; all links existing between the definitions and the syntactic cooccurrence of a lexeme must be made explicit; etc.
5. An ECD is a **formal** dictionary: all information is presented by means of a rich lexicographic metalanguage, which ensures a rigor never before attempted in lexicography.

Consistent with these five properties, an ECD features a standard, rigid structure for its entries, as described below.

3.2 Structure of an ECD Entry

Our discussion of ECD entries will draw on examples from the recently published Russian ECD (Mel'čuk and Zholkovsky 1984); the interested reader can also consult the first volume of the French ECD (Mel'čuk et al. 1984).

The basic unit of an explanatory combinatorial dictionary is a dictionary entry corresponding to a single **lexeme** or a single **phaseme** — one word or one phraseme taken in one separate sense. A family of dictionary entries for lexemes which are sufficiently close in meaning and which share the same signans (identical stem) is subsumed under one **vocable**, which is identified in upper-case letters.

Different vocables which are the same graphically (homonyms) are distinguished by numerical superscripts: Russ. BRAK¹ 'marriage' vs. BRAK² 'defective merchandise'.

The lexemes within a single vocable are distinguished by special indices mentioned here in order of increasing semantic proximity (cf. the use of numerical distinguishers for sememes, illustrated above):

Roman numerals differentiate lexemes whose definitions have a non-trivial common semantic component but whose semantic relationships are not sufficiently regular:

VYSOTA I 'height or altitude [as a vertical measure]'
 VYSOTA II 'level [of prices, aspirations, pressure]'
 VYSOTA III 'pitch [of sound]'
 VYSOTA IV 'altitude [of the sun over the horizon]'
 VYSOTA V 'altitude [of a triangle]'.

Arabic numerals indicate sufficiently regular semantic relationships among lexemes that share non-trivial semantic components, or regular polysemy: e.g. SPAT'¹ I.1 'sleep' vs. SPAT'¹ I.2 'be inert, as if sleeping'. The literal and the figurative meanings of **vspyxivat'** 'blaze up/flare up' are distinguished in this way, as are **kipet'** 'boil/seethe with', **zastyt'** 'congeal/become still [with fright, etc.]'; and the like.

Lower-case letters differentiate lexemes with slight semantic differences which are maximally regular: e.g. SKOROST' 1a 'speed' and SKOROST' 1b 'great speed'. The same goes for **vysota** 'height', **temperatura** 'temperature/high temperature', **kačestvo** 'quality/excellent quality', etc.

It should be stressed that all full idioms are entered in the ECD separately. (A **full idiom** is a phrase whose meaning cannot be computed from the meanings of its words and such that no constituent word retains its full meaning; cf. **shoot the breeze** or **More power to him!** in English.) Full idioms are provided with dictionary entries as if they were single lexemes; for example, **iz ljubvi k iskusstvu** 'for the love of the thing' or **kolot' glaza** 'to throw something in somebody's teeth'. Thus the ECD does not distinguish between mono-lexemic and multi-lexemic units as head entities in its entries.

An ECD entry is divided into three major zones (we will not even mention other subdivisions, which are less important in the present context, such as morphological information, usage labels, etc.):

- semantic zone
- syntactic zone
- lexical cooccurrence zone

3.2.1 Semantic Zone. The semantic zone contains the SemR of the entry lexeme, or its definition, written in accordance with the following six principles:

- (1) Generally speaking, the definiendum is not simply an entry lexeme but rather a **propositional form**: the lexeme with variables representing its semantic actants, or argument roles. Thus we define not 'to help' but rather 'X helps Y in Z by/with W'; not 'aggression' but

'aggression of X against Y'; not 'faithful' but 'X faithful to Y'; etc.

- (ii) The definiens is formulated in terms of lexemes which are semantically simpler than the entry lexeme (= definiendum). 'X is semantically simpler than Y' means that we need X to define Y but Y cannot (and will not) be used to define X. Thus if 'maneuvers of X' is defined as 'a large-scale training exercise of the armed forces X simulating combat', then 'maneuvers' cannot appear in the definitions of 'large-scale', 'train', 'exercise', 'armed forces', etc. This entails the **semantic decomposition** of any entry word, which in its turn, precludes vicious circles — that common plague of practically all existing dictionaries. Consistently applied, semantic decompositions lead to a set of semantic primitives.
- (iii) In the definiens, neither ambiguity nor synonymy of terms are allowed. To avoid ambiguity, the definition is written in terms of word senses, i.e. lexemes (lexical units supplied with distinctive indices, as specified above) rather than in terms of polysemous words. To avoid synonymy, just one expression is chosen and fixed explicitly for any meaning; e.g., we say only 'having the goal of' instead of '(in order) to', 'aiming at' or 'with the aim/goal of' etc. (In this way, an ECD ensures **univocality** of defining terms.)
- (iv) The definiendum and the definiens are strictly synonymous, and **ABSOLUTE MUTUAL SUBSTITUTABILITY** is required of them in all possible contexts. This means that a definition in an ECD contains only necessary components such that, taken together, they are sufficient to uniquely specify the definiendum. (This is the **adequacy** principle.)
- (v) Any lexemes that belong to the same vocable, i.e. that are semantically related, must show this relatedness explicitly — by displaying an identical semantic component, called **semantic bridge**.
- (vi) Any vocables that are semantically related must be organized according to the same pattern: they show roughly the same number and order of related lexemes, the definitions of lexemes are written following the same general schema, etc. (This is the **uniformity** principle; it is directly linked to the systematic character of an ECD and its field structure.)

As an example, let us quote the definition of Russ. SKOROST' la 'speed':

- (19) SKOROST' P-a/X-a - Y 'the speed of P/X being
 Y' = 'quantity II which characterizes 2 the motion
 I.1 P of X by the distance I.2a Y which X covers
 III in unit 4 time 2, or the value 6 of this
 quantity II'.

3.2.2 Syntactic Zone. An ECD fully specifies all the expressions which can be syntactically headed by the entry lexeme and which are "bound" by it (which are not free but rather idiomatic,) to a lesser or greater extent, with respect to the given entry. This is done by means of the so-called **government pattern (GP)**. A GP is a table in which each column represents one semantic actant of the lexeme (marked by the corresponding variable), and each element in the column represents one of the possible surface realizations of the corresponding syntactic actant. For instance, the GP for the Russian lexeme **skorost'** la 'speed' has the form:

1 = P or 1 = X	2 = Y
1. N _{gen}	1. Num _{nom} N' (v 'per' N'' _{acc})
2. A _{poss}	2. v' of' Num _{acc} N'(v 'per' N'' _{acc})
	3. Num _{nom} N
	4. v 'of' Num _{acc} N
	5. A
	6. N _{gen}

This GP specifies that the first deep-syntactic actant of the lexeme **skorost'** la 'speed' fills the slot of the variable P or X in its definition (1 = P or 1 = X, where X stands for the body, and P stands for the motion of the body whose speed is described). The second deep-syntactic actant fills the slot of the variable Y (2 = Y, Y being the value of the speed). The first deep-syntactic actant may be expressed on the surface either as a noun in the genitive case or as a possessive adjective (**skorost'** **samolëta** 'the speed of a plane', **naša skorost'** 'our speed'). The second deep-syntactic actant has six surface realizations:

- 1) a complex noun phrase in the nominative consisting of two phrases: a noun phrase "Numeral in the nominative + Noun denoting unit distance" and a prepositional phrase "v 'per' + Noun denoting unit time", for

starred example is followed by the number of the restriction it violates.

3.2.3 Lexical Cooccurrence Zone. This zone embodies the major novelty proposed in an ECD: namely, exhaustive and systematic description of restricted lexical cooccurrence of the entry lexeme. We mean the type of lexical collocation boldfaced in (20) — a dozen sentences collected in ten minutes from one newspaper page:

- (20)
- a. The President **clamped (imposed)** an overnight curfew on three areas ... **to stamp out (put down)** violence.
 - b. The panel **issued a report** to the Secretary of State¹.
 - c. President Reagan **rejected pleas** to open talks with striking US controllers.
 - d. Pope **released from hospital** [headline]. The Pope **left the hospital** yesterday, three months and one day after he was **struck by two bullets**. He **said a brief prayer** ...
 - e. The **heaviest prison terms** in Kentucky history (more than 1,600 years each) **have been handed down** against two men.
 - f. South African troops **have spread a dragnet** across the country in search for three **heavily armed** black guerillas.
 - g. The ANC **has claimed responsibility** for the **attack launched** last Tuesday in which four **rockets were fired** at an army camp.
 - h. We are looking for senior consultants of **proven competence** to **satisfy the demands** of our growing business.

Texts, from colloquial to artistic to technical, swarm with expressions of this type.

To describe all such collocations, the important concept of **lexical function (LF)** has been introduced.

A lexical function f is, like any mathematical function, a dependency that associates with a given "independent quantity" (the **argument**), a "dependent quantity" (the **value**). More precisely, an LF f associates with a lexical unit W (a word or a phrase) a set $\{W_1\}$ of (more or less synonymous) lexical units that express, contingent on W , a specific idea (such as 'very', 'begin', 'implement') represented by f . For example, the LF **Magn** (for the present glossed roughly as 'very') in

conjunction with the Russian words **naprjaženie** 'voltage' or **temperatura** 'temperature' is expressed by the adjective **vysokij** 'high'. In conjunction with **vyсота** 'height', however, the same function is expressed by **značitel'nyj** 'considerable', **bol'šoj** 'great' or **ogromnyj** 'enormous'; and in conjunction with **vibracija** 'vibration', by **sil'nyj** 'strong' or **intensivnyj** 'intense'. Thus we get **vysoкое naprjaženie** 'high voltage', **vyсокая температура** 'high temperature', **značitel'naja** (**bol'šaja**, **ogromnaja**) **vyсота** 'considerable (great, enormous) height', and **sil'naja** (**intensivnaja**) **vibracija** 'strong (intense) vibration'; but we do not get ***sil'noe** (**značitel'noe**) **naprjaženie**, ***bol'šaja** (**intensivnaja**, **sil'naja**) **temperatura**, ***intensivnaja** (**sil'naja**) **vyсота**, ***vyсокая** (**bol'šaja**) **vibracija**, etc.

The importance of LFs consists in the discovery of the following fact: in all natural languages there is only a limited number of meanings (about several dozen) that resemble 'very' in that they also each determine an LF.

And now, a formal definition of lexical function:

A dependency f is called lexical function if and only if it associates with a lexical unit W (a lexeme or a phraseme, the argument of f), a set $f(W) = \{W_1\}$ of lexical units (the value of f), such that the following two conditions are simultaneously met:

For any two different W^1 and W^2 , if $f(W^1)$ and $f(W^2)$ both exist, then:

1. Both $f(W^1)$ and $f(W^2)$ bear an identical relationship with respect to meaning and deep-syntactic role to W^1 and W^2 , respectively. [This condition is language-independent.]
2. In some cases, at least, $f(W^1) \neq f(W^2)$. [This condition is completely language-dependent; it means that in the given language, the value of f is **phraseologically bound** by its argument.]

An important proviso: A lexical function is not a genuine semantic unit, let alone a semantic primitive. LF's are introduced to describe restricted lexical cooccurrence and derivation, but by no means semantics. First, there are LF's that are semantically empty, their values being limited to purely syntactic roles (cf. **Oper_i**, **Func_i**, **Labor_i** below). Second, the expressions making up the value of a given LF f for a given argument need not be perfectly synonymous; it suffices for them to share a rather general and abstract meaning 'f' while differing in other components.

We will be interested in a particular type of LF, namely **standard lexical functions**, which form a proper subset of all lexical functions. A standard LF simultaneously satisfies two additional conditions:

3. The LF f is defined for a sufficiently large number of arguments. In other words, f has a sufficiently large semantic cooccurrence: its meaning ' f ' is sufficiently abstract to be compatible with a large number of other meanings. [This condition is language-independent.]
4. The LF f has a sufficiently large number of linguistic expressions as its possible values. In other words, the set of all $f(W_1)$, for a vast variety of W_1 , is sufficiently rich. [This condition is completely language-dependent.]

Let us illustrate this point with two examples. The meaning 'manufactured from very dark rye flour' in Russian has three expressions: *černyj* 'black', *ržanoj* 'rye [Adj]' and *iz ržanoj muki* 'from rye flour', and these expressions are phraseologically bound: Only *xleb* 'bread' or *suxar* 'rusk' can be called *černyj*, while very dark rye *buločka* 'bun', *bublik* 'bagel', *blin* 'pancake', *korž(ik)* '(a kind of) flat, dense, dry bread', *le pščka* 'a flat cake', etc. cannot; **černaja buločka* 'black bun' is readily understandable but ungrammatical.

Only *xleb* 'bread' and *lepščka* 'flat cake' but nothing else can be called *ržanoj*. (Notice that *černyj xleb* is always *ržanoj xleb*, but *ržanoj xleb* can be *svetlyj* 'light', as well as *černyj*.)

(Speaking of *buločki*, *bubliki*, *bliny*, *korži* and *koržiki*, Russian uses *iz ržanoj muki* (but not with reference to *suxari* 'rusks' and hardly with reference to *xleb*): **černyj/ržanoj bublik*, etc.)

As we see, this meaning satisfies Conditions 1 and 2 (*černyj: xleb = iz ržanoj muki: bublik = ržanaja: lepščka* vs. **ržanaja buločka*, ...) and therefore specifies a lexical function. However, this meaning violates Conditions 3 and 4: it is conceivable only with the names of bread-like baked products made from dough (so that it is semantically too specific) and it has only three expressions. Therefore, it is not a standard LF.

The meaning 'it is necessary to P this X' has in Russian a phraseologically bound expression *nuždat'sja* 'need' that is possible with some P's while impossible with others: *Ėta stat'ja nuždaetsja v ispravlenii <v dorabotke>* 'This paper needs to be corrected <more work>' but not **Ėta stat'ja nuždaetsja v soxranenii <v otpravke avtoru>* 'This paper needs to be saved <to be sent to the author>'. This meaning satisfies, in addition to Conditions 1 and 2, Condition 3 (it is extremely abstract), but it still violates Condition 4. (There are not numerous synonymous expressions. In fact, there is only one other expression for the meaning in question satisfying Condition 1: *trebovat* 'require', which is in free variation with *nuždat'sja*.)

Both above-mentioned meanings, 'manufactured from very dark rye flour' and 'it is necessary to P this X', are non-standard LF's in Russian. In what follows, only standard LF's are considered.

Since both defining conditions for standard LF's include the vague expression 'sufficiently large number', there is no sharp borderline between standard and non-standard LF's. That is, there are no formal criteria to tell them apart. This fact reflects the graduality so typical of natural languages. The concept of standard LF is fuzzy, as are most linguistic concepts.

Within the class of standard lexical functions, we will distinguish **simple** LF's and **compound** LF's, the latter being built out of the former. Notice that **simple** in this context by no means implies 'elementary' or 'further unanalyzable': some of our simple LF's could be represented in terms of other simple LF's (and thus could be treated as compound). Nevertheless, for purely linguistic reasons (primarily, frequency of occurrence) we consider a particular set of LF's as simple and take this set to be the basis of lexicographic descriptions.

The list of lexical functions which comprises the main body of our systematic survey is found in Appendix A. It includes only simple standard LF's. (This allows us to omit the adjectives "simple standard" everywhere, since the omission cannot lead to a confusion.)

Along with the LF's listed in Appendix A, two further types of LF's are extensively used in the Explanatory Combinatorial Dictionary: non-standard and compound LF's.

A **non-standard** LF is a meaning that is idiomatically expressed depending on a key word, but that has either a strongly limited semantic combinability or a fairly limited range of expressions, or both. In other words, it is too specific, too particular to be granted the status of a standard LF. Non-standard LF's are written in standardized natural language. Some examples:

- (21) such that Y is confined to his home (arest 'arrest') = domašnij 'house-[arest']
- (22) such that it is the result of a loss at cards that was not immediately paid (dolg 'debt') = kartočnyj 'card' [dolg], obsolete: [dolg] čestj '[debt] of honor'
- (23) during a short time and/or nonintensively (with the purpose of knowing Y somewhat better) (učit'sja 'learn') = // podučit'sja 'learn a bit (of something)'

A **compound** LF is a combination of syntactically related simple LF's that has a unique lexical expression covering the meaning of the combination as a whole. I have presented numerous examples of compound LF's in Appendix A; let me give some more illustrations, with the key word printed in boldface:

- AntiMagn** : židkie aplodismenty 'thin (lit. 'liquid') applause'
slabye dovody 'weak arguments'
nizkaja temperatura 'low temperature'
neznačitel'nye poteri 'negligible losses'
- AntiVer** : ložnyj styd 'false shame'
lživoe obeščanie 'false (lit. 'lying') promise'
ošibočnoe predstavlenie 'a wrong conception'
bezosnovatel'nye opaseniia 'unfounded misgivings/
 fears'
- IncepOper₁** : priobretat' populjarnost' 'acquire popularity'
vpadat' v otčajanie 'sink into despair'
vstavat' na put' predatel'stva 'take the path of
 treason'
perexodit' v pike 'go into a dive [as of an
 aircraft]'
- CausOper₂** : sdavat' v ekspluataciju 'put into operation'
vvergat' v rabstvo 'plunge into slavery'
stavit' pod kontrol 'put under control'
- AntiReal₂** : provalit'sja na ekzamene 'fail an examination'
otvergat' sovet 'reject a piece of advice'
otklonjat' xodatajstvo 'turn down an application'

The following four remarks bearing on all LF's are in order:

(i) An LF may have a **fused** expression, i.e. a lexical unit that does not include the key word but covers both the meaning of the function itself and that of its argument (the key word). The fusion is shown by the symbol // separating all the fused values (on its right) from all the non-fused values. For example:

Magn(dožd' 'rain') = prolivnoj 'heavy' // liven' 'shower' [i.e., liven' = prolivnoj dožd'; cf. Engl. downpour = heavy rain].

Magn(vkusno 'delicious') = očen' 'very' // pal'čiki obližeš', lit. 'You'll lick your fingers' [pal'čiki obliže' = očen' vkusno]

(ii) Several LF's which simultaneously have the same key word but are syntactically not linked to one another may be expressed by one lexical unit covering the meanings of all the LF's involved. This is what we call **configuration** of LF's (as opposed to compound LF's, in which all the constituent-simple LF's are syntactically linked). In a configuration of LF's, the "+" sign is used to separate the constituents. For example, in the entry SUD'BA 1 'fate, destiny' the notation

Fact₃^{II} + AntiBon₂: presledovat' 'persecute'

means that fate really affects [= **Fact**₃^{II}] the person in question (the defendant, so to speak) and its verdict is bad [= **AntiBon**₂]. Two further examples:

A₁(vosxiščenie 'delight') + **Magn**(vosxiščenie) = preispolnennyj
[vosxiščenijsja] 'full [of delight]'

Oper₁(otčajanie 'despair') +

Magn(otčajanie) = byt' vo vlasti [otčajanijsja] 'be completely in
[despair]'

(iii) Some LF's (most often, **Magn** or **Real**₁ with the latter's relatives, **Fact**₁ and **Labreal**₁) may be subscripted with a semantic component of the key-word's definition (in square brackets) to indicate that the meaning of this LF interacts with exactly this component of the key-word's meaning. Thus:

Labreal₁₂[xranit' 'keep'] (pamjat' 'computer memory') = xranit' [v pamjati] 'store [in memory]'

Labreal₁₂[vydavat' 'output'] (pamjat') = izvlekat' [iz pamjati] 'extract
[from the memory]'

Magn[bojat'sja 'be afraid'] (strax¹ 'fear') = dikijs 'wild', žutkijs
'terrible'

Magn[terjat' samokontrol' ... 'lose ... self-control'] (strax¹) =
paničeskijs 'panic [adj.]', životnyjs 'animal'

AntiVer[poražat' 'hit'] (streljat'¹ 'shoot') = ploxo 'badly', skverno
'poorly'

AntiVer[cel' 'target'] (streljat'¹) = v vozdux 'into the air'

(iv) Furthermore, some LF's may be superscripted with semantic labels, like 'usual', 'loc(ation)', 'temp(oral)', 'quant(itative)', to make their meaning more precise:

Magn^{temp}(opyt 'experience') = dlitel'nyjs 'long'

Magn^{quant}(opyt 'experience') = bol'šoj 'considerable'

In concluding this survey of LF's we would like once again to call attention to the fact that they are used for two main purposes:

- 1) for the description of idiomatic or restricted lexical cooccurrence or derivational relations; and
- 2) for specifying universal synonymic transformations of utterances on the deep-syntactic level.

It is clear, for example, that in any language (24) holds:

$$(24) W = \text{Oper}_1 \xrightarrow{-2} S_0(W) = \text{Oper}_2 \xrightarrow{-2} S_0(W) \\ = \text{Func}_1 \xrightarrow{-1} S_0(W) = \text{Func}_2 \xrightarrow{-1} S_0(W) = \text{Labor}_{12} \xrightarrow{-3} S_0(W); \text{ and so on.}$$

[Numbers on the arrows stand for deep-syntactic actants; e.g., the key word of an *Oper* is its second deep actant, etc.]

The transformations presented in (24) can be exemplified as follows:

- (25) vlijat' '(to) influence' = okazyvat' [= *Oper*₁] vlijanie [= *S*₀]
 'have
 influence' = byt', naxodit'sja [= *Oper*₂] pod vlijaniem 'be under
 influence', etc.

Compare (26):

- (26) Ivan durno vlijaet na Petra
 'John influences Peter badly' =

Ivan okazyvaet na Petra durnoe vlijanie
 'John has a bad influence on Peter' =

Pëtr naxoditsja pod durnym vlijaniem Ivana
 'Peter is under the bad influence of John'.

Rules of type (24) allow one to develop a paraphrasing system for synonymic transformations of sentences and/or discourses. Such a system can automatically produce, for any given text, a set of its synonymous or nearly-synonymous paraphrases. It also can automatically derive, for a set of synonymous texts, a canonical invariant.

Appendix A: List of Lexical Functions

A lexical function, notated by a boldfaced Latin abbreviation, is written before the parentheses enclosing its argument (a familiar mathematical functional notation). For example:

Function Argument	Function Value
Magn (shave)	= close, clean
Magn (easy)	= very, extremely, ... as pie
Magn (scoundrel)	= unmitigated, of the first water
Magn (condemn)	= strongly, in strongest terms
Magn (cold)	= very, terribly, ... enough to freeze the balls off a brass monkey
(Magn is from Lat. magnus 'great, big'.)	

As stated in Section 3.2.3, LF's are one of the central notions to a new type of dictionary, the Explanatory Combinatorial Dictionary. The simple standard lexical functions will be listed below in the order in which they appear in a lexical entry in such a dictionary.

1. **Syn** - synonym; **Syn**_⊃, **Syn**_⊂, and **Syn**_∩ designate, respectively, synonyms with broader, with narrower, and with intersecting meanings. (Symbols ⊃, ⊂, and ∩ have the same meaning when used with **Conv**, **Anti** and other LF's.) Examples: **Syn**(streljat 'shoot') = palit 'fire'; **Syn**(streljat 'shoot') = obstrelivat 'fire upon, shell, machine-gun'; etc.

2. **Conv**_{ij} - converseive, i.e., a lexical item with the same meaning as the key word **W** but with deep syntactic actants **i** and **j** permuted: **Conv**₂₁(inključat 'include') = prinadležat 'belong to [a set]'; **Conv**₂₃₁(mnenie 'opinion') = reputacija 'reputation'. ['Reputation', in contrast to 'opinion', is necessarily held by several people; this is why it is a narrower converseive.]

3. **Anti** - antonym: **Anti**(pobeda 'victory') = poraženje 'defeat'.

4. **Gener** - Generic concept such that '**Gener** + **W**' = '**W**' (where **W** is the key word): **Gener**(gaz 'gas') = veščestvo 'substance' [cf. gazoobraznoe veščestvo 'gaslike substance' = gaz 'gas']

5. **Figur** - Standard metaphor for **W**: **Figur**(blokada 'blockade') = kol'co, lit. 'ring' [kol'co blokady 'the grip of a blockade; a siege']; **Figur**(tuman 'fog') = pelena 'curtain' [pelena tumana 'curtain of fog'].

6-9. **S**₀, **A**₀, **Adv**₀, **V**₀ - syntactic derivatives of **W**; that is: noun (= substantival), adjective, adverb, and verb, respectively, which have the same meaning as **W**. Examples: **S**₀(streljat 'shoot') = strel'ba 'shooting'; **A**₀(streljat 'shoot') = strelkovyj 'shooting [attrib.]'; etc.

10. S_1 - standard name of the i -th participant in the situation described by W : $S_1(\text{učit' 'teach'}) = \text{učitel' 'teacher'}$; $S_2(\text{učit' 'teach'}) = \text{učenik 'pupil'}$; $S_3(\text{učit' 'teach'}) = (\text{učebnyj}) \text{ predmet ' (subject) matter [as in high school]}'$.

11-15. S_{instr} , S_{med} , S_{mod} , S_{loc} , S_{res} - standard name of instrument, means, mode, location, and result of the situation described by W : $S_{instr}(\text{streljat' 'shoot'}) = \text{ognestrel'noe oružie 'firearm'}$; $S_{med}(\text{streljat' 'shoot'}) = \text{boepripasy 'ammunition'}$; $S_{mod}(\text{rassmatrivat' 'consider'}) = \text{vzgljad [na čto-l.] 'a view [of something]'}$; $\text{podxod [k čemu-l.] 'approach [to something]'}$; $S_{loc}(\text{sražat'sja 'fight [as of two armies]'}) = \text{pole bitvy/boja 'battlefield'}$; $S_{res}(\text{učit'sja 'learn'}) = \text{navyki 'skills', znaniya 'knowledge'}$.

16. $Sing$ - 'one instance/unit of': $Sing(\text{gorox 'peas'}) = \text{gorošina 'pea'}$; $Sing(\text{celovat' '[to] kiss'}) = \text{pocelovat' 'give a kiss'}$.

17. $Mult$ - 'aggregate of': $Mult(\text{korabl' 'ship'}) = \text{flot 'fleet'}$; $Mult(\text{student 'student'}) = \text{studentčestvo 'student body'}$.⁶

18. Cap - 'head of': $Cap(\text{universitet 'university'}) = \text{rektor 'president'}$; $Cap(\text{fakul'tet 'faculty, school'}) = \text{dekan 'dean'}$.

19. $Equip$ - 'staff/crew of': $Equip(\text{teatr 'theatre'}) = \text{truppa 'troupe'}$; $Equip(\text{bol'nica 'hospital'}) = \text{personal 'personnel'}$; $Equip(\text{brak' 'marriage'}) = \text{suprug 'spouses'}$.

20. $Centr$ - 'center/culmination of': $Centr(\text{les 'forest'}) = \text{čašča [lesa] 'the thick [of the forest]'}$; $Centr(\text{slava 'glory'}) = \text{veršina [slavy] 'summit [of glory]'}$; $Centr(\text{bor'ba 'struggle'}) = \text{apogej [por'by] 'climax [of struggle]'}$. $Centr$ is current in combination with Loc_{in} (see below): $Loc_{in}Centr(\text{pustynja 'desert'}) = \text{v serdce [pustyni] 'in the heart [of the desert]'}$; $Loc_{in}Centr(\text{doroga 'road'}) = \text{posredi [dorogi] 'in the middle [of the road]'}$.

21. A_1 - determining property of the i -th participant of a situation characterizing him according to his role in the situation: $A_1(\text{gnev 'anger'}) = \text{v [gneve] 'in [anger]'}$, $\text{razgnevannyj 'angry'}$; $A_1(\text{slězy 'tears'}) = \text{v [slezax] 'in [tears]'}$; $A_1(\text{skorost' 'speed'}) = \text{so [skorost'ju...] 'with a speed of...'} [compare \text{spusk s takoj skorost'ju 'the descent with such a speed'}]$; $A_2(\text{streljat' 'shoot'}) = \text{pod obstrelom 'under fire'}$.

22. $Able_1$ - determining property of the i -th potential participant of a situation ('such that it can ... easily'/'such that it can be ... easily'): $Able_1(\text{plakat' 'cry'}) = \text{slezlivyj 'tearful'}$; $Able_2(\text{somnevatsja 'doubt'}) = \text{somnitel'nyj 'doubtful'}$.

23. $Magn$ - 'very', 'to a (very) high degree': $Magn(\text{temperatura 'temperature'}) = \text{vysokaja 'high'}$; $Magn(\text{rassmatrivat' 'examine'}) = \text{vnimatel'no 'attentively', pristal'no 'fixedly, intently'}$.

24-25. **Plus**, **Minus** - respectively, 'more' or 'less' (or 'to a greater/lesser extent') [than something else].

26-27. **Plus**^{refl}, **Minus**^{refl} - indicate that the comparison is made with a former state of the same object: **IncepPredPlus**^{refl}(temperatura 'temperature') = povyšat'sja 'rise, increase' [For the Lf's **Incep** and **Pred**, see below.]

28. **Ver** - 'as it should be' [meeting intended requirements]: **Ver**(udivlenie 'surprise') = iskrennee 'sincere', nepoddel'noe 'unfeigned'; **Ver**(sosud 'container') = celyj 'whole', germetičnyj 'hermetical, leak-proof'; **Ver**(pribor 'instrument') = točnyj 'precise'.

29. **Bon** - 'good' [a standard praise for W]: **Bon**(rezat 'cut') = akkuratno 'neatly, cleanly'; **Bon**(sudno¹ 'ship') = komfortabel'noe 'comfortable'.

30. **Pos**₁ - a standard praise of one of the participants of the situation denoted by W [but not of the situation itself]: **Pos**₂(recenzija 'review') = položitel'naja 'positive', while **Bon**(recenzija 'review') = xorošaja 'good', zamečatel'naja 'excellent', ..., blestjaščaja 'brilliant'; however, a blestjaščaja recenzija 'a brilliant(ly written) review' may well be položitel'naja 'positive' or otricatel'naja 'negative', that is, **Pos**₂ or **AntiPos**₂.

[NB: The Lf's **Magn**, **Ver**, **Bon** and **Pos**₁ are often combined with **Anti**. Thus, for instance, **Magn**(temperatura 'temperature') = vysokaja 'high', and **AntiMagn**(temperatura 'temperature') = nizkaja 'low'; **Pos**₂(mnenie 'opinion') = položitel'noe 'positive', and **AntiPos**₂(mnenie 'opinion') = otricatel'noe 'negative'.]

31. **Adv**₁ - determining property of an action by the i-th participant of a situation according to his role in the situation: **Adv**₁(slězy 'tears') = so [slezami] 'with [tears]'; **Adv**₁(skorost 'speed') = so skorost'ju ... 'at a speed of ...' [cf. mčat'sja so skorost'ju ... 'tear along at a ... speed']; **Adv**₂(somnevatsja 'doubt') = vrjad li 'hardly'.

32-34. **Loc**_{in}, **Loc**_{ab}, **Loc**_{ad} - preposition governing W [= the name of the situation] and designating a type of localization in space with the respective meaning — position, moving away, moving toward. Examples: **Loc**_{in}(vysota 'height') = na [vysote] 'at [a height]'; **Loc**_{ad}(vysota 'height') = na [vysotu] 'to [a height]'; **Loc**_{ab}(vysota 'height') = s [vysoty] 'from [a height]'.

35. **Loc**_{in}^{temp} - a preposition [analogous to **Loc**] with the meaning of temporal localization: **Loc**_{in}^{temp}(arest 'arrest') = pri [areste] 'while being [arrested]'; **Loc**_{in}^{temp}(analiz 'analysis') = v xode [analiza] 'in the course of [analysis]'.

36. **Instr** - a preposition [analogous to **Loc**] with the meaning of instrumentality: **Instr**(pistolet 'pistol') = iz [pistoleta], lit. 'with [a pistol]'; **Instr**(mašinka 'typewriter') = na [mašinke] 'on [a typewriter]'.

37. **Propt** - a preposition with the meaning 'because of', 'as the result of': **Propt**(strax 'fear') = ot [straxa], so [straxu] 'from [fear]'; **Propt**(ljubov 'love') = iz [ljubvi k ...] 'because of [one's love of ...]'; **Propt**(opyt I.1 'experience') = na [svoëm opyte] 'from [one's own experience]'.

38. **Copul** - a copula: **Copul**(učitel 'teacher') = byt', rabotat' [učitelem] 'be, work as [a teacher]'; **Copul**(primer 'example') = byt', javiťat'sja, služit' [primerom] 'be, represent, serve as [an example]'.

39. **Pred** - a verb meaning 'be W', i.e. semantically covering the syntactic combination of a **Copul**(W) with W. Thus **Pred** is nothing but a 'fused' expression of **Copul**(W) + W (on 'fused' expressions see below) needed for the convenience of some synonymic transformations. For example, **Pred**(pjanica 'drunkard') = pjanstvovat' 'drink heavily', **Pred**(rjadom 'next to') = sosedstvovat' '(to) neighbor'.

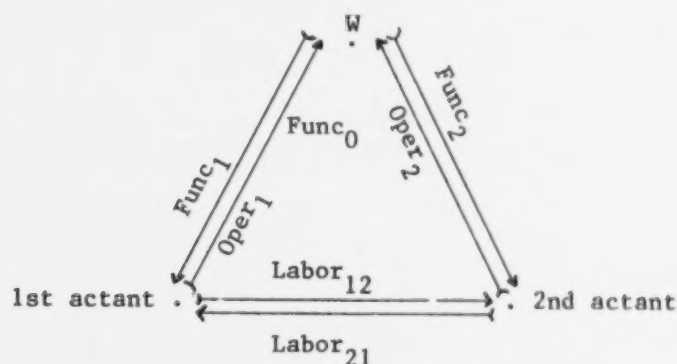
The next three LF's are verbs which are semantically empty in the context of the entry lexeme (i.e. their key word) and which serve to link, on the syntactic level, the name of a participant of a situation to W - the name of the situation itself. They play important semantico-syntactic roles and can be loosely called semi-auxiliaries.

40. **Oper**₁ - the first deep actant (and the surface subject) of this verb is the i-th participant of the situation, and the second deep actant (or the first surface object) of **Oper**₁ is W (further actants, if any, designate further participants of the situation); **Oper**₁(slězy 'tears') = lit', prolivat' 'shed'; **Oper**₁(arest 'arrest') = proizvodit' 'make [an arrest]'; **Oper**₂(arest 'arrest') = popadat' [pod arest] 'fall [under arrest]', podvergat'sja [arestu] 'undergo [arrest]'; **Oper**₁(soprotivlenie 'resistance') = okazyvat' 'show, put up'; **Oper**₂(soprotivlenie 'resistance') = vstrečat' 'meet', nataľkivat'sja [na soprotivlenie] 'run [into resistance]'.

41. **Func**₁ - the first deep actant (and the surface subject) of this verb is W - the name of the situation, and the second deep actant (and the first surface object), the i-th participant of the situation: **Func**₁(udivlenie 'surprise, astonishment') = oxvatyvati', lit. 'seize' [i.e. the person is overcome by surprise, astonishment]; **Func**₂(temperatura 'temperature') = ravnjat'sja 'be equal to'; **Func**₁(predloženie 'proposal') = isxodit' [ot kogo-l.] 'stem from, come from [someone]'; **Func**₂(predloženie 'proposal') = kasat'sja [čego-l.] 'concern [something]'. If there is no complement at all, i.e. **Func** is an intransitive verb, the subscript 0 is used: **Func**₀(dožd' 'rain') = idti, lit. 'walk' [cf. Engl. fall]

42. **Labor_{1j}** - the first deep actant (and the surface subject) of the verb is the *i*-th participant of the situation; the second deep actant, the *j*-th participant of the situation; and the third deep actant (implemented by the second surface object) is **W** itself. Examples: **Labor₁₂**(dopros 'interrogation') = podvergat' [kogo-1. doprosu] 'subject [someone to an interrogation]'; **Labor₃₂**(arenda 'lease') = sdavat' [čto-1. v arendu] 'grant [something on lease]'.

Oper, **Func**, and **Labor** can be paired in converse relations, that is, **Oper₁** = **Conv₂₁**(**Func₁**); **Labor₁₂** = **Conv₁₃₂**(**Oper₁**), and so on. These relationships may be diagrammed as follows:



Here, a two-participant situation designated by key word **W** is presented. Arrows represent semi-auxiliary verbs; the arrow's tail indicates the surface (grammatical) subject, the head pointing toward the first surface object.

A different way to express the same idea is by using a matrix:

Surface- Semi- auxiliary verbs	Surface subject	First surface object	Second surface object
Oper_{1/2}	1st/2nd deep actant of W	W	---
Func_{0/1/2/}	W	none/1st/2nd deep actant of W	---
Labor_{12/21}	1st/2nd deep actant of W	2nd/1st deep actant of W	W

$\text{PermOper}_2(\text{žkzamen 'exam'}) = \text{dopuskat' [kogo-1. k žkzamenu]}$ 'allow [someone to (take) an exam]'.

49. **Liqu** - 'liquidate', 'do something so that a situation does not occur or stops occurring': $\text{LiquFunc}_0(\text{negramotnost' 'illiteracy'}) = \text{pokončit' [s negramotnost'ju]}$ 'wipe out [illiteracy]'; $\text{Liqu}_1\text{Func}_0(\text{kostžr 'campfire'}) = \text{potužit' [kostžr]}$ 'extinguish [a campfire]'.

50-52. Now let us look at another triple of interrelated LF's: Real_1 , Fact_1 and Labreal_{1j} .

The LF's Real_1 , Fact_1 , and Labreal_{1j} are syntactically analogous to the functions Oper_1 , Func_1 and Labor_{1j} , respectively. This means that the names of the situation and of its 1-th participant fulfill with respect, for example, to Real_1 the same syntactic roles as they do with respect to Oper_1 ; etc. However, unlike the "empty" LF's Oper_1 , Func_1 , and Labor_{1j} , the lexical functions which we are now concerned with correspond to a specific meaning: 'fulfill a demand or requirement of ...'. The demands can differ for different W's. For example, the fulfillment, or realization, of a hypothesis is its confirmation; therefore, $\text{Real}_2(\text{gipoteza 'hypothesis'}) = \text{podverždat' 'confirm'}$ [$\text{Fakty podtverždajut gipotezu}$ 'The facts confirm the hypothesis'], and $\text{Fact}_2(\text{gipoteza 'hypothesis'}) = \text{sootvetstvovat' 'be in accordance with'}$ [$\text{Gipoteza sootvetsvuet faktam}$ 'The hypothesis is in accordance with the facts']. Realization of an artifact is its utilization according to its intended function; therefore, $\text{Fact}_0(\text{nož 'knife'}) = \text{rezat' 'cut'}$ [$\text{Etot nož režet xorošo}$ 'This knife cuts well']'. Further examples:

$\text{Real}_1(\text{obvinenie 'accusation'}) = \text{dokazyvat' [obvinenie]}$ 'prove [an accusation]'; $\text{Real}_1(\text{učebnoe zavedenie 'educational institution'}) = \text{prepodavat' [v učebnom zavedenii]}$ 'teach [in an educational institution]'.

$\text{Real}_2(\text{obvinenie}) = \text{soglašat'sja [s obvineniem]}$ 'agree [with an accusation]'; $\text{Real}_2(\text{učebnoe zavedenie}) = \text{izučat'sja, prepodavat'sja [v učebnom zavedenii]}$ 'be studied, be taught [in an educational institution]' (while $\text{Real}_3(\text{učebnoe zavedenie}) = \text{učit'sja [v učebnom zavedenii]}$ 'study [in an educational institution]'); $\text{Real}_2(\text{soblazn 'temptation'}) = \text{poddavat'sja [soblaznu]}$ 'yield [to temptation]'.

$\text{Fact}_0(\text{somnenie 'doubt'}) = \text{podtverždat'sja 'be confirmed'}$, $\text{opravdyvat'sja 'prove justified'}$; $\text{Fact}_0(\text{nadežda 'hope'}) = \text{sbyvat'sja 'come true'}$; $\text{Fact}_0(\text{sudno 'ship'}) = \text{plyt' 'sail'}$.

$\text{Fact}_1(\text{očered' 'turn'}) = \text{byt' [za kem-1.] 'be [someone's (turn)]'}$ [Očered' za vami 'It's your turn']; $\text{Fact}_1(\text{žksperiment 'experiment'}) = \text{udavat'sja [komu-1.] 'work out [for someone]'}$.

$\text{Fact}_2(\text{sudno}^1 \text{ 'ship'}) = \text{vezti, perevozit' [gruzy, passažirov]}$ 'convey, transport [cargo, passengers]'; $\text{Fact}_2(\text{sosud 'container'}) = \text{soderžat' [žto-1.] 'contain [something]'}$.

Labreal₁₂(viselica 'gallows') = vzdernut' [kogo-1. na viselicu] 'string up [someone on a gallows]'; Labreal₁₂(obed 'dinner') = est' [čto-1. na obed] 'eat [something for dinner]'.

The LF's Real₁, Fact₁, and Labreal₁₁ can be superscripted with Roman numerals to indicate the degree of the realization or fulfillment: the superscript I means fulfillment only at the psychological level, with the superscript II meaning fulfillment at the physical level, cf. Real₂^I(priglašenje 'invitation') = prinimat' 'accept', while Real₂^{II}(priglašenje) = sledovat' 'follow'; or Fact₁^I(čuvstvo 'emotion') = govorit' 'podskazyvat' 'tell', while Fact₁^{II}(čuvstvo) = zastavljat' 'force'.

53. Manif - 'manifest itself', 'become apparent': Manif(vina 'guilt', 'fault') = obnaruživat'sja 'become apparent'; Manif(udivlenie 'amazement') = skvožit' 'lurk'; Manif(bezgramotnost 'ignorance') = projavljat'sja 'manifest itself'.

Sympt - 'symptom', i.e. a verbal expression denoting a bodily reaction that is the symptom of an emotional or physical function: Sympt(udivlenie 'amazement', glaza 'eyes') = [U nego] glaza na lob polezli '[His] eyes started from his face'; Sympt(udivlenie 'amazement', rot 'mouth') = razinut' rot 'open [one's] mouth wide'; Sympt(strax 'fear', volosy 'hair') = [U nego] volosy vstali dybom '[His] hair stood on end'.

55. Prepar - 'prepare', 'get (something) ready for normal use or functioning': Prepar^IFact₀(revolver 'gun') = zarjažat' 'load'; Prepar^{II}Fact₀(revolver) = vzvodit' kurok 'raise the cock'. Roman superscripts, in much the same manner as with Real₁, Fact₁, and Labreal₁₁, express the degree of readiness. Compare also: PreparOper₁(obed 'dinner') = vyxodit' [k obedu] 'appear [for dinner]'; PreparOper₂(obed) = podavat' [na obed] 'serve [something for dinner]'; PreparFunc₁(obed) = podavat' [obed komu-1.] 'serve [somebody dinner]'.

56. Prox - 'be about to / on the verge of': ProxOper(otčajanie 'despair') = byt' na grani [otčajanija] 'be on the edge of [despair]'; Proxfunc₀(groza 'thunderstorm') = sobirat'sja 'gather, brew'.

57. Degrad - 'degradation', 'become worse or bad': Degrad(moloko 'milk') = skisnut' 'go sour'; Degrad(mjaso 'meat') = isportit'sja, protuxnut' 'go bad'; Degrad(disciplina 'discipline') = rasšatat'sja 'decay'.

58. Son - 'emit characteristic sound': Son(sobaka 'dog') = lajat' 'bark'; Son(banknoty 'banknotes') = xrustet' 'rustle'; Son(sneg 'snow') = skripet' 'crunch'; Son(vodopad 'waterfall') = revet' 'roar'.

59. Imper - 'do W!': Imper(streljat 'shoot') = ogon! 'fire!'; Imper(brat oružie 'seize arms') = v ruž'ě! 'take up arms!';

Imper(govorit' tixo 'speak lowly') = ts-s-s!, tš-š-š! 'sh-h-h';
Imper(brat' 'take') = na (te)! 'take it!', 'here!'.

60. **Perf** - 'perfective', i.e. 'have the process carried through to its natural limit': **Perf**(vstavat' 'be standing up') = vstat' 'have stood up'; **Perf**(rešat' 'be solving [a problem]') = rešit' 'have [it] solved'.

61. **Imperf** - 'imperfective', i.e. 'be carrying out the process': **Imperf**(vstat') = vstavat'; **Imperf**(rešit') = rešat'.

62. **Result** - 'resultative', i.e. 'the state of affairs that normally results from the completion of the process': **Result**(pokupat' 'buy') = imet' 'have'; **Result**(ložit'sja 'lay down') = ležat' 'be lying'; **Result**(naučit'sja 'have learnt') = umet' 'know', 'have necessary skills'.

Appendix B: Three sample dictionary entries

In order to give the reader the full taste of an ECD, we will quote two dictionary entries, one Russian and one French (from Mel'čuk and Zholkovsky 1984: 651-653 and Mel'čuk et al. 1984: 148-149, respectively). These are the lexemes **predloženie**¹ 1a,b and **proposition**¹ 1, both meaning roughly 'proposal' = 'the act of proposing or what is being proposed' and being approximate equivalents under translation.

The entries are reproduced here with slight modifications aimed at facilitating comparison. Since we address ourselves to an English-speaking audience, both actual metalanguages (Russian and French) have been replaced by English. We are not in a position to explain all the details of presentation; but we hope they will become clear in due course.

For the benefit of English readers, we add also an English lexical entry written by James Steele (Carleton University, Ottawa) and heavily edited by the author with the help of David Weber and David Tuggy (Summer Institute of Linguistics, Grand Forks, North Dakota). It is the first sense of the noun **hope** II (**hope** I being the verb).

SAMPLE 1: RUSSIAN

Russ. **PREDLOZENIE**/E¹, ja, ja, neuter

1a,1b. Predloženie X-a [Z-u] delat' Y 'Proposal by X [to Z] to do Y' = 'What is (being) proposed by X to Z' [= S₂(predlagat' 1a,1b or predložit' 1a,1b)].

Government Pattern

1 = X	2 = Y	3 = Z
1. N _{gen}	1. N _{gen}	
2. <u>ot</u> N _{gen}	2. V _{inf}	_____
3. A _{poss}	3. <u>ščtoby</u> + PROPOS	
4. A		

1) C_{1.2}: only with LF Result₃ and Conv₂₁Result₃

2) $C_{1.4}$: $A = A_0(N)$, where N is the name of a country

3) $C_{2.1}$: only for $PREDLOŽENIE^1$ la; N is a predicate

4) $C_{1.1} + C_{2.1}$: impossible

C_1 : predloženie Petra <Francii> 'proposal by Peter <France>', [Ja imeju] predloženie ot direktora 'I have] a proposal from the director', naši <francuzskie> predloženiya 'our <French> proposals'

C_2 : predloženie mira <pomošči> 'offer of peace <help>', predloženie otdoxnut' <čtoby my otdoxnuli> 'proposal to rest <that we have a rest>'

$C_1 + C_2$: americanskoe predloženie pomošči 'American offer of help', predloženie brigadira otdoxnut' <čtoby my otdoxnuli> 'the proposal by the brigade chief to rest <that we have a rest>'

Impossible: *Menja tronulo predloženie pomošči ot Koli 'I was moved by the offer to help from Kolya (1) [= Menja tronulo Kolino predloženie pomošči; but cf. U menja est' [= Conv₂₁Result₃] predloženie pomošči ot Koli 'I have an offer of help from Kolya']; *moskovskoe predloženie 'Moscow proposal' (2) [= predloženie Moskuy]; *predloženie otdyxa <knigi> 'proposal/offer of rest <of a book> (3); *predloženie Peti pomošči 'offer of help by Pete' (4)

Lexical Functions

V_0	: <u>predlagat' la,b, predložit' la,b</u> 'propose, offer'
Magn	: <u>nastojčivo</u> 'insistent'
Magnquant	: <u>neodnokratnye</u> 'multiple' P. is in pl
AntiMagn	: <u>robkoe</u> 'timid', <u>ostorožnoe</u> 'cautious'
Ver ₁	: <u>ser'ěznoe</u> 'serious'
AntiVer ₁	: <u>legkomyslennoe</u> 'flippant', <u>neobdumannoe</u> 'unthoughtful'
Bon ₁	: <u>ljubeznoe</u> 'amiable'
Bon ₂	: <u>vygodnoe</u> 'advantageous'
Magn + AntiBon ₁	: <u>nazojlivo</u> 'importunate'
Oper ₁ (P. ² → become the wife of X)	: <u>delat'</u> 'make' [-e N _{dat}]
F_1 = Oper ₁ +the same as P. ¹ of W	: <u>prisoedinjat'sja</u> 'join' [k ~ju N _{gen} = W] either $C_1 \neq \Delta$, or P. has an identifying modifier
PerfIncepOper ₃	: <u>polučit'</u> 'receive' [-e (ot 'from' N _{gen})]
ContFunc ₀	: <u>ostava'sja v sile</u> 'still hold'
Liqu ₁ Func ₀	: <u>brat' obratno</u> 'take back' [(svož) ~e]
Func ₁	: <u>isxodut'</u> 'come' [ot 'from' N _{gen}]
PerfFunc ₁	: <u>offic posledovat'</u> 'follow' [so storony

Func₂ : 'from the side of 'N_{gen}] | P. cannot have a Magn
 : sostojat', saključat sja 'consist'
 [v 'in' N_{prep}]
 ...

Examples

Ona odgovorila otkazom na ego robkoe predloženie pomoč' ej 'She
 answered with a refusal to his shy proposal to help her'. Po
 predloženiju xozjajki my ostalis' eščë na dva dnja 'Following the
 proposal of the lady of the house, we stayed two more days'.

SAMPLE 2: FRENCH

Fr. PROPOSITION¹, feminine

1. Proposition de X à Z de Ye r 'Proposal by X to Z to
 do Y' = 'Fact that X proposes Y to Z or the message concerning Y
 by means of which X proposes Y to Z' [= S₀/S₂(proposer 1)]

Government Pattern

1 = X	2 = Y	3 = Z
1. <u>de</u> N	1. <u>de</u> N	1. <u>à</u> N
2. A _{poss}	2. <u>de</u> V _{inf}	
3. A	3. <u>que</u> PROP _{subj}	

1) C_{1.3} : A = A₀(N), where N is the name of a country, or N =
patron, gouvernement, ...

2) C_{2.3} : M₁ = M₁(Y)

3) C₃ without C₁ or C₂ : not desirable

- C₁ : la proposition de Pierre <de la France>
 'proposal by Peter <France>, notre proposition
 'our proposal', les propositions françaises
 'French proposals'
- C₂ : la proposition de paix <d'aide> 'proposal
 of peace <help>', la proposition d'apporter de
l'aide 'proposal to give some help', la
proposition que son valet fasse cela 'proposal
 that his servant do that'
- C₁ + C₂ + C₃ : sa proposition <la proposition française> de
travail à Marie 'his <the French> proposal of
 work to Mary', sa proposition <la proposition
 française> à Marie de participer au tournoi 'his

<the French> proposal to Mary to participate at the tournament', la proposition d'argent de Pierre à Marie 'proposal of money by Peter to Mary'

Not desirable: ?la proposition à Marie (3) [= la proposition faite à Marie 'the proposal made to Mary']

Lexical Functions

Syn _C	: <u>offre</u> 'offer'
V ₀	: <u>proposer</u> 1 'propose'
Ver ₁	: <u>sérieuse</u> 'serious'
AntiVer ₁	: <u>à la légère</u> 'lighthearted'
Bon ₁	: <u>amicale</u> 'friendly'
Bon ₂	: <u>avantageuse</u> 'advantageous', <u>intéressante</u> 'interesting'
Oper ₁	: <u>faire</u> 'make' [ART ~]
ProxOper ₁	: <u>avoir</u> 'have' [ART ~ (<u>à</u> Oper ₁)] [J'ai une proposition (<u>à faire</u>) 'I have a proposition (to make)']
ContOper ₁	: <u>maintenir</u> 'maintain' [ART ~]
Oper ₃	: <u>recevoir</u> 'receive' [ART ~]
ResultOper ₃	: <u>avoir</u> 'have' [ART ~] [J'ai une proposition <u>avantageuse</u> 'I have an advantageous proposal']
ContFunc ₀	: <u>tenir encore</u> <toujours> 'still hold'
Liqu ₁ Func ₀	: <u>retirer</u> 'take back' [ART ~], <u>revenir</u> 'take back' [<u>sur</u> ART ~]
Func ₁	: <u>venir</u> 'come' [<u>de</u> N]
Func ₂	: <u>consister</u> 'consist' [<u>en</u> N]
Func ₃	: <u>s'adresser</u> 'be addressed' [<u>à</u> N]
...	

Examples

Bien qu'officiellement acceptée, ma proposition est restée lettre morte 'Although officially accepted, my proposal went unheeded'. Il accueillit ma proposition d'aide avec joie <froide> 'He received my proposal/offer of help joyfully <coldly>'.

SAMPLE 3: ENGLISH

II. HOPE, noun

1. X's hope of Y = S₀(hope I.1).

Government Pattern

1 = X	2 = Y
1. of N	1. for N
2. N's	2. of N
3. A _{poss}	3. of V
	4. to vger inf
	5. that PROP

N.B.: The plural [-s] can be used in the singular sense with C_{2.3,4} and, if N is singular, with C_{2.1,2} [His hopes <= hope> of winning the lottery prize died when his number was not drawn; He bought ten tickets for the lottery in hopes <= hope> of winning the prize].

- 1) C_{1.1} + C_{2.2} : not desirable
- C₁ : the hope of Peter, Peter's <his> hope
- C₂ : hope for <of> victory, hope of winning <to win> a victory, hope that victory would be won
- C₁ + C₂ : Peter's <his> hope of being successful <to be a success>, the hope of Peter <his hope> that success would be achieved
- Not desirable : ?the hope of Peter of success (1) = Peter's hope of success

Lexical Functions

Syn ₃	: desire; expectation	
Syn ₀	: wish, aspiration, dream; faith; prospects; confidence; optimism	
Contr	: fear	
Anti	: despair	
Anti ₀	: doubt; fear, dread	
Gener	: feeling, sense [of ~]	
Suddenly IncepFunc ₁ + Magn + Figur	: surge [of ~] X is a person; wave [of ~] X is many people	
IncepOper ₁ (surge of H.) F ₁ = Involv(wave of H.) ² →Z	: have, feel [a surge of ~]	H. in singular; C ₁ = A
CausF ₁	: sweep [across N = Z] : send [a wave of ~ across N] [President's speech sent a wave of hope across the country]	

Suddenly IncepFunc ₁ + AntiMagn + Figur	: flicker, glimmer, gleam, ray, spark [of ~]; breath [of ~]; grain, seed [of ~] ['H. is supposed to grow']	H. in singular; C ₁ = A
IncepOper ₁ (flash, glimmer, gleam, spark of H.)	: have [a flicker <glimmer, gleam, spark> of ~]	
IncepOper ₁ (ray of H.)	: see [a ray of ~]	
CausFunc ₁ (ray of H.)	: give [N = X a ray of ~] M ₁ is a fact	
AntiMagn + Figur	: shadow [of (a) ~] used in negative constructions, H. in singular [The search party continued to look for the lost child without a shadow of hope that the little creature would be found alive]	
Magn + A ₁	: full [of ~] H. in singular, filled [with ~]	
nonA ₁	: void, devoid [of ~], without [-]; lit or arch out [of ~] [The old king was out of hope when he learned of the disaster]	H. in singular; C ₁ = A
V ₀	: hope I.1	
Culm	: pinnacle [of ~]	
Loc _{in} Culm	: at the pinnacle [of ~]	
Magn['desire']	: big 1 [= 'intense'], strong <great 1 [= beyond the ordinary in strength'] [Their great hope for an early settlement did not materialize], high 1 [= 'intense to an extraordinary degree']; dear, earnest, passionate, profound, sincere; eager, lively <ardent, exultant, fervent, intense [However passionate <exultant, fervent> were their hopes for a coming change, they couldn't withstand the last blow]; unconquerable, irrepressible	
AntiMagn + A ₁ IncepPredPlus CausPredPlus	: budding : build up 2, bolster, encourage, expand 2, foster, increase 2, strengthen, spur, warm 2 [Ø/ART ~] [Do not allow an initial success to build up your hopes]; coll get up [A _{poss} ~] [Do not get your hopes up too soon] : [so, too ...] high, coll a lot : spread 1 [among N = X] [Hope spread among the starving villagers when rain clouds appeared on the horizon]	
Magn(Build up 1,2) IncepPredPlus ₁ quant		

CausPredPlusquant	: expand 2, increase 2, spread 2 [Ø/ART ~ among Ns] [The good news spread hope among the inhabitants]
IncepPredMinus	: diminish 1, dwindle, lessen 1, shrink 1, sink 1, weaken 1 [Their hopes diminished when they took this new consideration into account]; cool 1, cloud 1, darken 1, dim 1, fade, fail 1 [= 'become weaker'], go down, wane, wither 1 [Hopes faded after the defeat of their best regiment]
CausPredMinus	: diminish 2, lessen 2, shrink 2, sink 2, weaken 2, cool 2, cloud 2, dampen, darken 2, daunt, dim 2, wither 2, undermine, [Ø/ART ~] [The injury of their best player cooled their hope of winning the game]
Oper ₁	: entertain, have, hold out [Ø/ART ~]; experience, feel [Ø/ART ~]
AntiVer + Oper ₁	: indulge [ART ~] C ₂ ≠ Δ
IncepOper ₁	: find [-] C ₂ = Δ
(Plus +) IncepOper ₁	: gain [-]
ContOper ₁	: live [in ~]; live, rest [in the ~] C ₂ ≠ Δ [You must live in hope even though the odds are against you]
Caus ₁ ContOper ₁	: cling [to the ~] C ₂ ≠ Δ [= nonPerm ₁ FinOper ₁]
FinOper ₁	: lose [Ø/ART ~] // lose heart
CausOper ₁	: inspire [N with Ø/ART ~]
Oper ₁	: Not much - < No ~ (left) H. in singular, C ₁ = Δ
LiquOper ₁	: cheat [N of ART ~]
Liqu ₁ Oper ₁	: abandon, discard, give up, leave (behind), relinquish, surrender [Ø/ART ~]
A ₁ nonPermOper ₁	: past, beyond [-] H. in singular, C ₁ = Δ [The patient seemed beyond hope]
Able ₂ Oper ₁	: of [-] H. in singular [a child of great hope; a country of hope]
Oper ₂	: be [To visit Japan was his long-time hope]
Magn['probability']	: certain C ₂ ≠ Δ, confident, firm; lit and arch absolute G ≠ Oper ₁
AntiMagn['probability']	: doubtful, faint, feeble, forlorn 1 [= 'weak'], little, scant, slim, small, unsure; all G(H.) = abandon, lose [Abandon all hope!]
expressed cautiously + AntiMagn['probability']	: cautious [The cautious hope of the Minister of Finance was that new fis- cal measures would bring an end to the

Magn ₁	recession]
	: boundless, unbounded
Magn ₁ quant	: common, general < universal
AntiMagn ₁	: humble
Bon ₂ + Magn ₂	: big 2, bright; great 2 [= 'pre-eminent'], high 2 [= 'of exalted importance'], monumental, utmost [To reach the moon in a single-stage rocket was their big hope; Everyone has such high hopes for what can be done in Africa]
Magn ^{temp}	: A ₁ ContFunc _{0/1}
Bon to X	: dear, good, pleasant, pleasing, sweet < fondest
Ver	: real, substantial
AntiVer	: blind, drunken, deluded, empty, fallacious, false, fantastic, fond, futile, idle, illusory, naive, mad, spurious, unbridled, unfounded, unrealistic, unsubstantiated; vain; fallible; lit airy, lit beguiling
Magn['desire'] + AntiVer:	wild
Adv ₁	: with [the ~] [He attended the ball with the hope of meeting her] // hopefully ₁
nonAdv ₁	: without [~] [The little children wandered in the dark forest without hope; He continued his work, but without hope of completing it]//hope- lessly 1
Propt	: out of [~ for N = Y], in [~ of N/V _{ger}], in [~s of V _{ger}] C ₂ ≠ A [She cherished his picture out of hope for his return; He attended the ball in hope of meeting her; She wrote to him every week in hopes of his returning]
IncepPredPlus	: build up 1, expand 1, go up, grow, increase 1, rise, warm 1 [The rescuers' hope grew <rose> when they heard that the storm was subsiding]
Conv ₂₁ (CausFunc ₁)	: ride [on N N denotes the cause of H. [Our hopes ride on his recent achievement]
in order to make X perform P, CausFunc ₁	: dangle [ART ~ before N = X] H. in singular
Func ₂	: be [that PROP < for V _{ger} /N>] N denotes an action [Our only hope was that if someone was there, he would be afraid of us]
F ₂ = the realization of X's H. depends on Z	: [X's ~] lies [on <in> N]

H. in
singular

	[Hassan's hopes now lay on this little ridge] hangs [on N = Z], hinges [on V _{ger} /N = Z] N denotes an action
Caus ₁ F ₂	: fix, lay, set, fasten, hang, pin [A _{poss} ~ to N = Z] attach [A _{poss} ~ to N = Z]
Caus ₁ (H. is based on Z)	: base, build, rest [ART ~ on N = Z]
Real ₁	: fulfill, realize, reach, attain [ART ~]
Real ₂	: satisfy, meet [ART ~], meet [with ART ~]
Fact ₀	: come true, materialize, be realized
AntiFact ₀	: come to nothing, fail 3 [= 'prove misleading'] [The farmer's hope came to nothing when he was unable to develop a rust-resistant species of wheat]
CausFact ₀	: advance, further [Ø/ART ~]
nonPermFact ₀	: disappoint, frustrate [Ø/ART ~]
Magn + A ₁ Caus ₁ Manif	: bright, radiant, starry-eyed, wide-eyed [with ~] H. in singular
Caus ₁ Manif	: display, exude, show [Ø/ART ~]
by saying, Caus ₁ Manif	: express [Ø/ART ~]
nonPerm ₁ Manif	: hide, conceal [ART ~]
A ₂ nonPerm ₁ Manif	: secret, silent
Adv ₁ Caus ₁ Manif ² nonFunc ₀	: // hopelessly 2 [He shook his head hopelessly]
Excess ^{trem} (body)-Sympt ₁₂	: tremble [with ~]
Excessfulg(eyes	H. in singular
face)-Sympt ₁₂	: beam, glow [with ~]
CausDegrad	: corrupt, spoil [ART ~]
situation P being such that there is no H. that P will change for the better	: // hopelessly 3 P [He is hopelessly ill]
realize that Y of X's H. may take place later than X expected	: defer [Ø/ART ~] [Hope deferred makes the heart sick]
Caus ₁ ContFunc ₀	: cherish, harbour, nurse, sustain 1 [= 'keep hoping continuously'] [ART ~] C ₂ ≠ Δ [their generation nursed the hope of a promised land]
Caus ₁ ContFunc ₀ is a spontaneous human characteristic	: ~ springs eternal
FinFunc ₀	: collapse, die, disappear, end 1, fail 2 [= 'cease'], perish, vanish [Hope for a

	lasting peace collapsed with the renewed fighting]
PerfFinFunc ₀	: be gone, be finished, be extinct, be at an end; be dead [Our hope for life on Mars is dead]; lit or arch be done
S ₁ CausFunc ₀	: source [of ~]
FigurS ₁ CausFunc ₀	: star, beacon, well-spring, bastion [of ~] H. in singular
being caused by W, IncepFunc _{0/1}	: arise, be born [in N < N's soul, heart, breast>]
ContFunc _{0/1}	: linger, persist, remain [in N/among Ns]
A ₁ ContFunc _{0/1}	: enduring, lasting, persistent, steadfast, bibl unfading
CausContFunc _{0/1}	: feed, keep alive, nourish, sustain 2 [= 'give strength to hopes'] [Good news from the front fed the hopes of the general]
LiquFunc _{0/1}	: crush, demolish, destroy, wreck [Ø/ART ~]; cut off [A _{poss} ~/N from ~]; blast, blight, break, dash, end 2, extinguish, kill, ruin, shatter, take away, trample down [ART ~] [The mother's hopes were shattered by the death of her young son]; bury, drown [Ø/ART ~] M ₁ denotes a fact
Func ₁	: there be [in N's soul <heart, breast, ...>]
IncepFunc ₁	: spring [in N's soul <heart, breast, ... >]
again IncepFunc ₁	: return [to N], revive 1 [in N] [Hope revived in her]
nonPermContFunc ₁	: leave [N little <no> ~] [It leaves him no hope of escape]
FinFunc ₁	: abandon [N]
CausFunc ₁	: arouse, awaken [-]; raise [The serious negotiations raise hopes for an early end to the strike]; stir up [Ø/ART ~ in N]; engender, instill, plant [Ø/ART ~ in N] H. in singular; give, hold out, offer [Ø/ART ~ to N] H. in singular
again CausFunc ₁	: restore, revive 2, reawaken, rekindle [Ø/ART ~] [Warm weather restored the hope of the farmers]
live disregarding adverse facts or material necessities as if the H. of X	

constituted reality
or a means of X's
sustenance

: live [on ~] [Despite a fatal diagnosis,
the young man continued to live on hope;
After all food was consumed, the explor-
ers lived on hope]

Examples

Man is a victim of dope, in the incurable form of hope. (O.Nash) It was evident to everyone that a strong sense of hope was the guiding power in her life. A light of hope shone in their eyes when they heard the good news. Rising popularity in pre-election polls gave the new party a ray of hope. The discovery of penicillin opened a door of hope for the wounded. The children were full of hope after visiting Santa Claus. The young man was devoid of hope when he learned the sad fate of his brother. She was suddenly killed while at the pinnacle of hope. To feed the poor was the dear hope of the good monk. An ardent hope of victory is what encouraged the competitors to do their best. Inspired by certain hope of success, the hunter continued his search. They had but faint hope that the buried miners would be found alive. The children had boundless hopes that they would excel in the examination. The common hope was for a bountiful harvest. Their best hopes were unmercifully shattered by the death of their son. Throughout their sinful lives, they were inspired by the sweet hope of redemption. Only substantial hopes should be taken into consideration by the committee. He was always preoccupied with empty hopes. Driven to extremes by her wild hopes, she finally committed suicide. When the soldier heard of the great victory at the front, his hopes for an end to the war went up. The woman's budding hopes grew stronger when she received the happy letter from her friend. The silly man indulged the hope that he could write a novel in three days. The sick man gained hope when he learned of the new medicine that might cure his illness. The workers found new hope in their revolution. The young child clung to the hope that her dead father would come back to life. Hope for a solution to the problem must be deferred until better conditions prevail. When the radio finally went silent, the operators in the control tower lost all hope. A happy turn of events inspired the men with hope. Stormy weather cheated the fishermen of their hope for a large catch. Our hopes now lie in the skill of the surgeon. The great drought left them little hope for a good crop. Once the oil slick was found on the surface of the water, all hope for the submarine was gone. Hope for a new life was born in John's soul after he read the Bible. Despite initial losses, hope for an early victory lingered. Winning the hand of a certain lady was the romantic gentleman's steadfast hope. The long drought destroyed the hope of the farmer for a good harvest. Lack of financial resources cut off his hopes. The young lover walked down the path, hope springing in his heart. After the war, hope for economic progress returned. The serious negotiations stirred hopes for an early end to the strike. Let us instil in our children hope for a better world. The physician could not hold out much hope to his

patient. The coach's hopes for his team rode on their recent achievement. The Minister of Housing dangled the hope of interest-free loans before builders. The young man pinned his hope (of quick success <for this marriage>) on finding steady employment. The newly elected Prime Minister was a source of hope to many citizens. Medicare was their beacon of hope. Do not hang your hopes on impossible schemes if you wish to avoid disappointment. Despite the dense fog, the explorers never gave up hope of finding their way back to their camp. The young bride clung to the hope that her spouse would return from the war alive. Unconquerable hope guided the romantic poet. In composing a great symphony, the musician realized a boyhood hope. The farmer's hopes came true after he developed a species of wheat that was resistant to rust. Development of the electrical automobile advanced the car manufacturers' hopes for success. Adverse economic conditions frustrated hopes of achieving a low rate of unemployment. She was radiant with hope after being nominated a candidate for office. An optimist will display hope even in adversity. The fisherman expressed hope for a good catch. The villain concealed a hope that the plane would crash with his mother on board. The silent hope of the innocent victim was her ultimate vindication. When the evangelist touched the sick woman, she trembled with hope. The children were glowing with hope on Christmas Eve. No adversity will corrupt our hope.

Notes

I am happy to express my heartfelt gratitude to Lidija Iordanskaja, who, as always, was the first to read the text of my paper and suggest many important corrections and improvements, and to Peter Constable, Ingrid Meyer and James Steele, who kindly accepted the task of revising the English of the manuscript.

A part of this paper was presented as key-note address at the 1985 annual meeting of the Linguistic Association of Canada and the United States (LACUS), in Saskatoon, Saskatchewan, Canada.

1. We make use of the distinction available in German:

sprachlich 'pertaining to or related to language',
sprachwissenschaftlich 'pertaining to or related
to linguistics'

We use the English word 'linguistics', therefore, only when the second meaning is intended, and the German sprachlich when the first meaning is intended.

2. The primacy of the speaker's viewpoint for linguistics is so fundamental an issue that it seems recommendable to analyze here an interesting argument against it, advanced by Thomas R. Hofmann. The argument runs approximately as follows:

Within the framework of "good" communication, the speaker has to make sure that he is properly understood; and to achieve

this, he must put himself in the addressee's shoes, that is, try to decode all his own utterances and see whether they can be interpreted the way he intends them to be. Therefore, the speaker must use listener's device as well, with the result that the task of speaking is harder than that of understanding.

The reasoning is quite sound -- but it misses the point. Certainly, the speaker has to use a listening (= understanding) device, but it by no means follows from this fact that the speaking device is more complex than the listening one. The task of a speaker in a communication act may be harder than that of his listener; we do not, however, compare communication tasks, but rather mechanisms correlating meanings and texts (leaving aside mechanisms for constructing meanings, for backtracking what has been said, etc. — mechanisms that are extremely important in "sprachlich" communication but do not pertain to language and are, consequently, of no interest for linguistics).

3. The logico-semantic concepts **predicate**, **place** (of a predicate) and **argument** are central to the present discussion, but we are in no position to introduce them here. We assume our readers' familiarity with them and limit ourselves to a warning against confusing **predicate** in logico-semantic sense (= 'a meaning having obligatory slots for other meanings') with the grammatical term **predicate** (= 'main verb'). Likewise, **place** (= 'a slot of a predicate') and **argument** (= 'meaning filling the slot of a predicate') should not be taken in their everyday sense.

4. For lack of space, we cannot even touch here upon the fundamental problem of semantic primitives.

5. Once again, we are not in a position to substantiate here the use of purely distinctive (= meaningless) symbols instead of meaningful labels on the arcs.

6. Interesting examples of values of the LF **Mult** in English can be found, most unexpectedly, in **The Book of Lists**, by D. Wallechinsky, I. Wallace and A. Wallace, New York: W. Morrow, 1977, page 135:

A murder	of crows
clowder	cats
leap	leopards
sloth	bears
raffer	turkey
smack	jellyfish
skulk	foxes
labor	moles
crash	rhinoceros
siege	herons
rag	colts
drift	hogs
charm	finches

trip	goats
knot	toads
parliament	owls
troop	kangaroos
gaggle	geese
pride	lions
muster	peacocks

"Although not frequently heard in conversation, these terms are fully correct and appropriate ways of describing the animal listed" (ibid.).

7. A very good illustration of values of the LF Real (and AntiReal) in English is offered in Montreal's **The Gazette**, Sept. 28, 1985, in a letter to Ann Landers, which we reproduce here:

Dear Ann:

I discovered this item stuck in an old book.

It was written in longhand and there is no sign of an author. I hope you will print it.

I found it highly inspirational.

- T. D.

Dear T.D.:

So did I. Here it is:

What is life?

Life is a challenge ... meet it.
 Life is a gift ... accept it.
 Life is an adventure ... dare it.
 Life is a sorrow ... overcome it.
 Life is a tragedy ... face it.
 Life is a duty ... perform it.
 Life is a game ... play it.
 Life is a mystery ... unfold it.
 Life is a song ... sing it.
 Life is an opportunity ... take it.
 Life is a journey ... complete it.
 Life is a promise ... fulfill it.
 Life is a beauty ... praise it.
 Life is a struggle ... fight it.
 Life is a goal ... achieve it.
 Life is a puzzle ... solve it.

References

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

- Chomsky, Noam. 1966. Topics in the theory of generative grammar. The Hague-Paris: Mouton.
- . 1968. Language and mind. New York etc.: Harcourt, Brace and Jovanovich.
- Lakoff, Robin. 1973. Lexicography and generative grammar. II: context and connotation in the dictionary. In: R. I. McDavid and A. R. Duckert (eds.), *Lexicography in English* [Annals of the New York Academy of Sciences, 211], New York, 154-164.
- Lamb, Sydney. 1966. Outline of Stratificational Grammar. Washington: Georgetown University Press.
- Mel'čuk, Igor A. 1981. Meaning-Text Models: A Recent Trend in Soviet Linguistics. *Annual Review of Anthropology*, 10, 27-62.
- , with Nadia Arbathevsky-Jumarie; Léo Elnitsky; Lidiya Iordanskaja; and Adèle Lessard. 1984. Dictionnaire explicatif et combinatoire du français contemporain. *Recherches lexico-sémantiques I*. Montréal: Les Presses de l'Université de Montréal.
- and Zholkovsky, Alexander K. 1984. *Tolkovo-kombinatornyj slovar' sovremennogo russkogo jazyka. Opyty semantiko-sintaksičeskogo opisanija russkoj leksiki* [Explanatory combinatorial dictionary of modern russian]. Vienna: Wiener Slawistischer Almanach.
- Nakhimovsky, Alexander. 1983. Meaning-text linguistics and the problem of voice. Carbondale-Edmonton: Linguistic Research.
- Sanders, Gerald A. 1972. Equational grammar. The Hague-Paris: Mouton.
- Sgall, Petr. 1967. *Generativní popis jazyka a čezká deklinace*. Praha: Academia.
- Testen, David; Veena Mishra; and Joseph Drogo (eds.) 1984. Papers from the parasession on lexical semantics. Chicago: Chicago Linguistic Society.
- Wierzbicka, Anna. 1972. *Semantic Primitives*. Frankfurt/M.: Athenäum.
- . 1980. *Lingua Mentalis: The Semantics of Natural Language*. Sydney etc.: Academic Press.
- Žolkovskij, Aleksandr K., and Igor A. Mel'čuk. 1965. O vozmožnom metode i instrumentax semantičeskogo sinteza [On a Possible Method and Instruments for Semantic Synthesis (of Texts)]. *Naučno-texničeskaja informacija*, No 6, 23-28.